Flock, H. R. Some Conditions Sufficient for Accurate Monocular Perceptions of Moving Surface Slants 560
Fowler, H. See Wischner, G. J.
Galanter, E. See Linker, E.
Garner, W. R., and Creelman, C. D. Effect of Redundancy and Duration on Absolute Judgments of Visual Stimuli 168
Gibson, E. J., Bishop, C. H., Schiff, W., and Smith, J. Comparison of Meaningfulness and Pronunciability as Grouping Principles in the Perception and Retention of Verbal Material 173
Ginsberg, R. Stimulus Overlap in a Massed-Trial Situation 553
Gogel, W. C. Perception of Depth from Binocular Disparity 379
Gold, C. See Mandler, G.
Goldfarb, J. L. See Goldstone, S.
Goldstein, S. See Bourne, L. E., Jr.
Goldstone, S., and Goldfarb, J. L. Direct Comparison of Auditory and Visual Durations .. 483
Goodrich, K. P. Effect of a Ready Signal on the Latency of Voluntary Responses in Eyelid Conditioning 496
Gormezano, I. See Abraham, F.
Grant, D. A. See Levy, C. M.
Greeno, J. G. Paired-Associate Learning with Massed and Distributed Repetitions of Items. 286
Grings, W. W. See Lockhart, R. A.
Guy, D. E. See Weinberg, D. R.
Haber, R. N. A Replication of Selective Attention and Coding in Visual Perception 402
Hadley, R. W. See Riley, D. A.
Hall, R. C. See Wischner, G. J.
Hamilton, C. R., and Bossom, J. Decay of Prism Aftereffects 148
Harrow, M. Stimulus Aspects Responsible for the Rapid Acquisition of Reversal Shifts in Concept Formation 330
Hawker, J. R. Effects of Prompting and Confirmation in a Serial Learning Task 99
Haygood, R. C., and Bourne, L. E., Jr. Forms of Relevant Stimulus Redundancy in Concept Identification 392
Helson, H., and Bevan, W. An Investigation of Variables in Judgments of Relative Area ... 335
Hermelin, B. M., and Venables, P. H. Reaction Time and Alpha Blocking in Normal and Severely Subnormal Subjects 365
Hillner, K. P. See Peterson, L. R.
Homzie, M. See Trapold, M. A.
Homzie, M. J. See Spence, K. W.
Horowitz, L. M., Lippman, L. G., Norman, S. A., and McConkie, G. W. Compound Stimuli in Paired-Associate Learning 132
Hulse, S. H., and Firestone, R. J. Mean Amount of Reinforcement and Instrumental Response Strength 417
Hunt, D. P. Effects of Nonlinear and Discrete Transformations of Feedback Information on Human Tracking Performance 486
Irion, A. L. See Koonce, J. M.
Ison, J. R. Acquisition and Reversal of a Spatial Response as a Function of Sucrose Concentration 495
Jeffrey, W. E., and Cohen, L. B. Effect of Spatial Separation of Stimulus, Response, and Reinforcement on Selective Learning in Children 577
Jeffrey, W. E. See Atkinson, R. C.
Jenkins, J. J., and Bailey, V. B. Cue Selection and Mediated Transfer in Paired-Associate Learning 101
Johnson, D. M., and King, C. R., Jr. Systematic Study of End Anchoring and Central Tendency of Judgment 501
Jones, A. Drive and Incentive Variables Associated with the Statistical Properties of Sequences of Stimuli 423
Jones, M. B. See Bilodeau, E. A.
Kaplan, I. T., and Metlay, W. Light Intensity and Binocular Rivalry 22
Kaplan, S. See Kleinsmith, L. J.
Kendler, H. H., and Woerner, M. Nonreinforcements of Perceptual and Mediating Responses in Concept Learning 591
Kimble, G. A. See Nicholls, M. F.
King, C. R., Jr. See Johnson, D. M.
Kleinsmith, L. J., and Kaplan, S. Interaction of Arousal and Recall Interval in Nonsense Syllable Paired-Associate Learning 124
Klemmer, E. T. Does Recoding from Binary to Octal Improve the Perception of Binary Patterns? 19
Koonce, J. M., Chambliss, D. J., and Irion, A. L. Long-Term Reminiscence in the Pursuit-Rotor Habit 498

Korn, J. H. See Ludvigson, H. W.
Korn, J. H. See Moyer, K. E.
Lakso, V. See Miller, M. E.
Larimer, G. S., and White, B. J. **Some Effects of Monetary Reward and Knowledge of Results on Judgment** 27
Laughlin, P. R. **Speed versus Minimum-Choice Instructions in Concept Attainment** 596
Lehr, D. J. See Bergum, B. O.
Levy, C. M., Grant, D. A., and Clark, A. H. **Reversal of Conditioned Discrimination of the Eyelid Response** 80
Levy, C. M. See Bilodeau, E. A.
Link, W. E. See Bourne, L. E., Jr.
Linker, E., Moore, M. E., and Galanter, E. **Taste Thresholds, Detection Models, and Disparate Results** 59
Lippman, L. G. See Horowitz, L. M.
Lockhart, R. A., and Grings, W. W. **Interstimulus Interval Effects in GSR Discrimination Conditioning** 209
Longstreth, L. E. **Partial Reinforcement Effect and Extinction as a Function of Frustration and Interfering Responses** 581
Ludvigson, H. W., Caul, W. F., Korn, J. H., and McHose, J. H. **Development and Attenuation of Delay-Engendered Avoidance Behavior** 405
McConkie, G. W. See Horowitz, L. M.
McHose, J. H. See Ludvigson, H. W.
McKee, J. P. See Riley, D. A.
MacKinnon, J. R., and Amsel, A. **Magnitude of the Frustration Effect as a Function of Confinement and Detention in the Frustrating Situation** 468
Mandler, G., Cowan, P. A., and Gold, C. **Concept Learning and Probability Matching** 514
Metlay, W. See Kaplan, I. T.
Miller, M. E., and Lakso, V. **Effect of Constant versus Varied Pairing of Simultaneous Intentional- and Incidental-Learning Materials with Different Rates and Numbers of Exposures** 256
Miller, R. E., and Murphy, J. V. **Influence of the Spatial Relationships between the Cue, Reward, and Response in Discrimination Learning** 120
Minucci, P. K., and Connors, M. M. **Reaction Time under Three Viewing Conditions: Binocular, Dominant Eye, and Nondominant Eye** 268
Moore, M. E. See Linker, E.
Morning, N., and Voss, J. F. **Mediation of R_1-R_2 in the Acquisition of S-R_1, S-R_2 Associations** 67
Moyer, K. E., and Korn, J. H. **Effect of UCS Intensity on the Acquisition and Extinction of an Avoidance Response** 352
Moylan, J. J. **Kinesthetic Figural Aftereffects: Satiation or Contrast** 83
Murphy, J. V. See Miller, R. E.
Newman, S. E. **A Replication of Paired-Associate Learning as a Function of S-R Similarity** 592
Nicholls, M. F., and Kimble, G. A. **Effect of Instructions upon Eyelid Conditioning** 400
Nieuwenhuyse, B. See Zajonc, R. B.
Norman, S. A. See Horowitz, L. M.
Passey, G. E., and Sekyra, F. **Geometric and Arithmetic Means as Indexes of UCS Intensity with Variable Reinforcement** 7
Pennypacker, H. S. **External Inhibition of the Conditioned Eyelid Reflex** 33
Peterson, C., and Ulehla, Z. J. **Uncertainty, Inference Difficulty, and Probability Learning** 523
Peterson, J. R. See Fitts, P. M.
Peterson, L. R., and Hillner, K. P. **Minimal Paired-Associate Learning** 300
Peterson, M. J. **Cue Trials, Frequency of Presentation, and Mediating Responses** 432
Petre, R. D. **Concept Acquisition as a Function of Stimulus-Equivalence Pretraining with Identical and Dissimilar Stimuli** 360
Pollio, H. R. **Composition of Associative Clusters** 199
Postman, L. **Acquisition and Retention of Consistent Associative Responses** 183
Price, L. E. See Vandament, W. E.
Prokasy, W. F., and Ebel, H. C. **GSR Conditioning and Sensitization as a Function of Intertrial Interval** 113
Puff, C. R. See Bousfield, W. A.
Rapoport, J. **Adaptation in the Perception of Rotary Motion** 263
Riley, D. A., McKee, J. P., and Hadley, R. W. **Prediction of Auditory Discrimination Learning and Transposition from Children's Auditory Ordering Ability** 324
Rock, I., and Brosgole, L. **Grouping Based on Phenomenal Proximity** 531
Rosner, B. S. **Temporal Interaction between Electrocutaneous Stimuli** 191
Rotberg, I. C. **Verbal Paired-Associate Learning as a Function of Grouping Similar Stimuli or Responses** 298
Rule, S. J. **Effect of Instructional Set on Responses to Complex Sounds** 215

RUTLEDGE, E. See TRAPOLD, M. A.
RUTLEDGE, E. F. See SPENCE, K. W.
SCHIFF, W. See GIBSON, E. J.
SCHILD, M. E. See BATTIG, W. F.
SEGAL, E. M. **Demonstration of Acquired Distinctiveness of Cues Using a Paired-Associate Learning Task** 587
SEKYRA, F. See PASSEY, G. E.
SHOEMAKER, R. See ATKINSON, R. C.
SIMON, S., AND WOOD, G. **Backward Learning and the Stimulus-Familiarization Inhibitory Effect** 310
SIMON, S. H. **Differential Prediction and Postdiction of Win-Lose Events in a Spatial-Generalization Problem** 342
SMITH, J. See GIBSON, E. J.
SOMMER, G. R. See ATKINSON, R. C.
SPEAR, N. E., EKSTRAND, B. R., AND UNDERWOOD, B. J. **Association by Contiguity** 151
SPENCE, J. T. **Verbal Discrimination Performance under Different Verbal Reinforcement Combinations** 195
SPENCE, K. W., HOMZIE, M. J., AND RUTLEDGE, E. F. **Extinction of the Human Eyelid CR as a Function of the Discriminability of the Change from Acquisition to Extinction** 545
SPERLING, S. E., AND VALLE, F. P. **Handling-Gentling as a Positive Secondary Reinforcer** 573
TEICHNER, W. H. **Temperature Acclimatization, Response Strength, and Thermal Preferences in the Rat** 221
THEIOS, J., AND BRELSFORD, J. **Overlearning-Extinction Effect as an Incentive Phenomenon** 463
TRABASSO, T., AND BOWER, G. **Presolution Reversal and Dimensional Shifts in Concept Identification** 398
TRABASSO, T. See CONANT, M. B.
TRAPOLD, M. A., HOMZIE, M., AND RUTLEDGE, E. **Backward Conditioning and UCR Latency** 387
TUPPER, R. W. See WEINBERG, D. R.
TURNER, E. D. See BEVAN, W.
UHL, C. N. **Effect of Overlapping Cues upon Discrimination Learning** 91
ULEHLA, Z. J. See PETERSON, C.
UNDERWOOD, B. J. See SPEAR, N. E.
UNGER, S. M. **Habituation of the Vasoconstrictive Orienting Reaction** 11
VALLE, F. P. See SPERLING, S. E.
VANDAMENT, W. E., AND PRICE, L. E. **Primary Stimulus Generalization under Different Percentages of Reinforcement in Eyelid Conditioning** 162
VENABLES, P. H. See HERMELIN, B. M.
VOSS, J. F. See MORNING, N.
WATKINS, W. H. **Effect of Certain Noises upon Detection of Visual Signals** 72
WEINBERG, D. R., GUY, D. E., AND TUPPER, R. W. **Variation of Postfeedback Interval in Simple Motor Learning** 98
WHITE, B. J. See LARIMER, G. S.
WIEHE, R. See ABRAHAM, F.
WIPF, J. L. **Positive and Negative Gradients of Response Strength in a Temporal Conflict Situation** 234
WISCHNER, G. J., HALL, R. C., AND FOWLER, H. **Discrimination Learning under Various Combinations of Food and Shock for "Correct" and "Incorrect" Responses** 48
WOERNER, M. See KENDLER, H. H.
WOOD, G. See SIMON, S.
YOUNG, R. K., AND CASEY, M. **Transfer from Serial to Paired-Associate Learning** 594
YOUNG, R. K., AND CLARK, J. **Compound-Stimulus Hypothesis in Serial Learning** 301
ZAJONC, R. B., AND NIEUWENHUYSE, B. **Relationship between Word Frequency and Recognition: Perceptual Process or Response Bias?** 276

Journal of
Experimental Psychology

VOL. 67, NO. 1 JANUARY 1964

TRANSFER OF VASOCONSTRICTION OVER A BIPOLAR MEANING DIMENSION

LOREN E. ACKER AND ALLAN E. EDWARDS

University of California, Los Angeles *Veterans Administration Center, Los Angeles*

Transfer of a salivary response conditioned to the word "good" to sentences which appeared to vary in meaning from good to bad has been recently reported. This phenomenon was replicated and extended by conditioning (24 Ss) vasoconstriction to the words "good" or "bad," testing for transfer on words previously rated by S on the evaluative scale of the semantic differential, and either instructing or not instructing Ss to covertly evaluate the test words. Transfer of vasoconstriction ($p < .001$) was demonstrated, but the semantic differential ratings were not significantly changed by the differential conditioning. It was concluded that transfer, as predicted by a mediation model, was demonstrated.

Razran, in a review of current psychological research in the Soviet Union, reported an experiment by V. D. Volkova (1953) on the exteroceptive conditioning of one pole of an evaluative dimension and the subsequent transfer of the conditioned response to sentences whose meaning varied along that dimension. Specifically, salivation to the word GOOD with a differentiation of no salivation to the word BAD was established with a classical conditioning procedure. Subsequent to this conditioning *S* was given what *E* considered nine good sentences, six bad sentences, and one neutral sentence interspersed with reconditioning and differentiation trials. Volkova found that salivation occurred to good sentences while little or no salivation occurred to bad sentences and an intermediate amount to the neutral sentence. In a similar experiment Volkova (1957) demonstrated transfer of salivary and motor reflexes over a "correct-mistake" dimension.

Volkova's experiment using the good-bad dimension of meaning needs several modifications in order to assess this transfer phenomenon more adequately: (*a*) an objective and quantitative measure of the goodness or badness of the stimulus to which the autonomic response may generalize; (*b*) a counterbalancing of the conditioned stimuli so that one may assess the reliability and generality of the phenomenon.

The present study is designed to

make the above modifications and also (*a*) to use vasoconstriction instead of salivation, providing a continuous recording of the response measure, (*b*) to determine the effects of experimentally varying instructions to *S* to evaluate subvocally the test words on the relevant meaning dimension, and (*c*) to use single words instead of sentences as the test stimuli, thus permitting an objective and quantitative index of the "goodness" or "badness" of these stimuli by the good-bad scale of the semantic differential (Osgood, Suci, & Tannenbaum, 1957).

The good-bad scale from the semantic differential was utilized not only in order to provide an index of meaning but also because Volkova's results could be interpreted in the context of mediation theory, i.e., salivation occurred to good sentences by means of the mediation provided by the meaning common to these sentences and the conditioned stimulus GOOD. In the same manner little salivation occurred to bad sentences since there was no meaning common to these sentences and the conditioned stimulus. Thus, in the present study generalization of vasoconstriction along a bipolar evaluation dimension of meaning will be studied.

METHOD

Apparatus and Materials

The good-bad scale from the semantic differential was centered and printed on 3 × 5 in. white cards. At the top of the scale and in the middle of each card appeared one word to be rated. These words were picked from normative data gathered by Jenkins, Russell, and Suci (1958). There were 75 words in all, 25 with normative ratings of 6 or 7 (bad), 25 with normative ratings of 4 (neutral), and 25 with normative ratings of 1 or 2 (good).

Either GOOD or BAD, for a given group, was the conditioned stimulus, and white noise lasting 1,000 msec. at 110 db. was the unconditioned stimulus. A program of differential conditioning with a CS-UCS interval of 500 msec., where the positive CS was reinforced 75% of the time, was transcribed onto a magnetic tape. A set of Koss earphones was used by *S* for receiving instructions and stimuli from *E*.

An Offner dynagraph was employed to amplify and record the pneumoplethysmographic ac signal from *S*'s left index finger.

Subjects

A total of 29 volunteer male patients from the gastrointestinal ward of a general medical and surgical hospital were used as *S*s (none of whom had hearing difficulties). Of the 29 *S*s used 5 did not show conditioning and thus were dropped from the experiment leaving a total *N* of 24.

Procedure

Each *S* was run individually to conclusion (about 2 hr.) without interruption.

Phase I.—The *S* rated the 75 words on the semantic differential cards. He was instructed to rate each word on the 7-point scale under it, not to hesitate making extreme ratings, leave the card blank if he did not understand the meaning of a word, and to rate as neutral any word that caused him to vacillate between a rating of bad and good.

After *S* had rated all the words he was taken into the dynagraph laboratory, placed in a supine position, and the plethysmograph pickup and earphones were affixed. The *S* was left to relax and settle down as *E* moved to an adjoining room where the recording and programing equipment were located. The *S*'s semantic differential cards were sorted into four groups of words rated 1, 7, 4, and residual. From each of the first three groups five words were picked at random plus one additional filler word that had been rated as neutral.

At this point *S*s were alternately assigned to one of two groups, an instructed evaluation group (G_{IE}) and a noninstructed evaluation group (G_{NIE}).

Phase II.—The *S* was now told that he would hear some words and that he should lie as still as possible. If *S* was in G_{NIE} he was told nothing more in this phase. However, if *S* was in G_{IE} he was told to evaluate each word, to himself and not out loud, as either "good," "bad," or "neutral" as soon as the word was presented. From the recording room *E* operated the dynagraph and tape recorder and administered the 15 selected words in a random order preceded by the 1 filler word, with an interstimulus interval of approximately 15 sec., simultaneously recording them for further use. The 16 words were presented twice in the same order.

Phase III.—At the conclusion of the above, all *S*s were instructed that they would now hear the words GOOD and BAD and they were to repeat each word as soon as they heard it. The words GOOD and BAD were presented four times each in a random order providing a base plethysmographic measure against which conditioning could later be measured. After these base measures were taken, *S* was told that he would now hear the words GOOD and BAD repeatedly along with some very loud noise. At this point the two groups of *S*s were subdivided again. Two groups (G_{IE-G}) and (G_{NIE-G}) were told that a loud noise would follow GOOD most of the time but not all of the time and would never follow BAD. The two other groups (G_{IE-B}) and (G_{NIE-B}) were similarly instructed interchanging GOOD and BAD. The *S*s in all groups were told to repeat the word GOOD or BAD as soon as they heard it. At this time the appropriate prerecorded program of GOOD and BAD plus white noise was presented. The program was repeated until *E* observed four consecutive large constrictions, on nonreinforced trials, of the plethysmograph to the conditioned stimulus word along with four consecutive little, or no, constrictions to the differentiated word.

Phase IV.—The tape made during Phase II was played to *S* immediately following Phase III. The *S*s in G_{NIE-G} and G_{NIE-B} were again told only that they would hear some words. The *S*s in G_{IE-G} and G_{IE-B} were again told to rate each word, to themselves, as being either good, bad, or neutral. The plethysmograph recording was continued.

After Phase IV, *S* was taken from the laboratory to an adjoining room where he rerated the original 75 words on the semantic differential cards.

RESULTS

A prestimulus measure from the plethysmograph was obtained by utilizing the two systolic spikes just prior to stimulus presentation. These two spikes were averaged in terms of their height in millimeters from a constant diastolic baseline. A poststimulus measure from the plethysmograph was obtained by averaging, in millimeters from the constant diastolic baseline, the three smallest adjacent spikes within 10 sec. after stimulus presentation. The difference between the pre- and the poststimulus measure for a given stimulus yielded a plethysmographic change (PC) score.

TABLE 1

MEAN VASOCONSTRICTION (μl/10 cc) BEFORE AND AFTER CONDITIONING FOR THREE WORD CATEGORIES

Words	Before Conditioning	After Conditioning
Maximal	.031	.062
Intermediate	.038	.040
Minimal	.040	.027

Note.—$N = 24$.

For a given *S* the PC scores for the expected maximal transfer words (words previously rated "good" for G_{IE-G} and G_{NIE-G} and words rated as "bad" for G_{IE-B} and G_{NIE-B}) from Phase II were averaged over both repetitions given before conditioning, and both repetitions given after conditioning. In the same manner PC scores for intermediate and expected minimal transfer words (words rated good for G_{IE-B} and G_{NIE-B} and words rated bad for G_{IE-G} and G_{NIE-G}) were averaged over both repetitions given before and after conditioning. Each *S* thus provided six mean PC scores, one for each of the three word categories before conditioning and one for each of the three word categories after conditioning. The mean PC scores were entered into a four-way analysis of variance (Lindquist, 1953) with CS (GOOD or BAD) and instructions (instructed evaluation or noninstructed evaluation) as the between-*S*s variables, and word categories (expected maximal, intermediate, and minimal

TABLE 2

MEAN VASOCONSTRICTION (μl/10 cc) BEFORE AND AFTER CONDITIONING FOR GOOD, NEUTRAL, AND BAD WORDS, BOTH CONDITIONED STIMULI

Words	Conditioned Stimulus GOOD		Conditioned Stimulus BAD	
	Before Conditioning	After Conditioning	Before Conditioning	After Conditioning
Good	.040	.070	.034	.018
Neutral	.037	.047	.040	.033
Bad	.044	.034	.021	.054

Note.—$N = 24$.

transfer) and before-after conditioning as the within-*S*s variables.

The pertinent test involves the Word Categories × Before-After interaction means, $F(2, 40) = 14.35$, $p < .001$, tabulated in Table 1. The obtained p demonstrates that conditioning had an effect upon the relationships among the words, and, more specifically, Duncan range tests between before and after conditioning for each of the three word categories indicate a significant change only for expected maximal transfer words ($p < .001$). Also, the intermediate and expected minimal transfer words differ from the expected maximal transfer words after conditioning, ($p < .02$, $p < .001$, respectively) but do not differ significantly from one another before conditioning.

The main effect of word categories is significant, $F(2, 40) = 11.08$, $p < .001$, but is uninterpretable due to the significant Word Categories × Before-After interaction effect. Instruction, and none of its interactions, did not approach significance.

To determine whether using GOOD or BAD as a CS affects the transfer phenomena, it is necessary to reevaluate the data without collapsing the good, neutral, and bad words into word categories. In this analysis, the Good, Neutral, and Bad Words × Before-After × CS interaction would be the pertinent point of interest, $F(2, 40) = 14.35$, $p < .001$. This interaction is shown in Table 2.

Duncan range ($p < .01$) tests indicated that maximal (i.e., good words for G_{IE-G} and G_{NIE-G} and the bad words for G_{IE-B} and G_{NIE-B}) words show a significant increase in constriction both for good words and for bad words. All other changes do not reach significance. Of the before-conditioning measures only the bad and neutral words when the CS is BAD differ significantly. When the CS is GOOD, the neutral and bad words both differ significantly from the good words after conditioning. When the CS is BAD, the analogous case exists. In neither case do the minimal and intermediate words differ significantly from one another.

A Lindquist Type I analysis of variance was performed on the semantic differential rating, defining the 75 words as good, (1–3), neutral, (4), or bad, (5–7), according to *S*'s first rating. A given *S*'s scores were his mean changes in these three categories, subtracting his second rating from his first rating. The resulting mean changes were entered into the analysis of variance with the categories, good, neutral, and bad, as the within-*S*s variable and the conditioned stimuli of GOOD and BAD as the between-*S*s variable. The category variable was significant ($p < .02$). All other sources failed to reach significance, the Category × Stimuli interaction having $p < .10$. The category effect reflected good words being rated less good and bad words being rated less bad on the second rating.

DISCUSSION

The generalization of vasoconstriction along a bipolar evaluative dimension of meaning by all four groups of *S*s confirms Volkova's research. The conception of an evaluative meaning dimension, as defined by ratings on the good-bad bipolar scale, predicted accurately the words to which transfer of a conditioned vasoconstriction would or would not occur. In the case of G_{NIE-G}, for example, vasoconstriction conditioned to the word GOOD transfers to a word mediated by the "good" meaning of that word. The conditioned response does not transfer to words rated as neutral or bad because these words lack the appropriate mediator. The argument is analogous in the case of each of the other groups.

In the case of instructed as against noninstructed evaluation, a variable not considered by Volkova's study, the results failed to show any differential transfer, i.e., instructed evaluation or the lack of it fails to alter the transfer of the conditioned response. Perhaps the inconsequential effect of *E*'s instructions to subvocally evaluate the words could in part be due to all *S*s acquiring a set or habit to evaluate all words as either good, bad, or neutral regardless of instructions by having rated the 75 words before the test session.

Branka (1957) has shown that expectation of the UCS upon presentation of the generalization stimulus was a necessary condition for generalization. In the present study expectation of the white noise upon presentation of the expected maximal transfer words should have had a heightened probability of occurrence in G_{IE-G} and G_{IE-B}. Here *S* was actually required to subvocally emit the CS. Since, in this study, no independent measure of white noise expectancy was obtained, Branka's results cannot be seriously qualified by the lack of differential results between the instructed and noninstructed groups.

Making available, or blocking, the mediators necessary for semantic generalization has been shown by Hartman (1963), and Lovibond (1959), under quite dissimilar experimental conditions to be inconsequential to subsequent generalization. This is in agreement with the results of the present study if one interprets the manipulation of awareness by instructions as a manipulation of the availability of the mediator necessary for the generalization of the vasoconstriction response.

The changes noted for the semantic differential ratings are difficult to interpret. Possibly a regression effect accounts in part for the tendency for good words to be rated less good and bad words to be rated less bad on the second ratings. It had been thought that perhaps the pairing of the noxious white noise with GOOD or BAD in the conditioning phase would differentially affect subsequent semantic differential ratings when compared to ratings made prior to the conditioning experience. The lack of any significant interaction between changes in ratings and the conditioned stimulus utilized does not bear this out.

As seen in Tables 1 and 2 there is a parallelism between good and neutral words, in terms of before- and after-conditioning scores, regardless of which CS was used. Although the neutral words never show significant transfer when the good words do, the parallelism warrants further study as an indicant of a possible semantic predisposition to categorize neutrality as good when the only alternative is bad.

A similar replication of Volkova's (1957) study of transfer over a correct-mistake dimension of meaning is suggested. Such a replication, with appropriate controls and objectification, could greatly enhance the generality of the transfer phenomenon of autonomic responses over dimensions of meaning.

REFERENCES

BRANKA, A. A. Semantic generalization at the level of the conditioning experiment. *Amer. J. Psychol.*, 1957, **70**, 541–549.

HARTMAN, T. F. Semantic transfer of the differential eyelid response from words to objects. *J. exp. Psychol.*, 1963, **65**, 194–200.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles for 360 words. *Amer. J. Psychol.*, 1958, **71**, 688–699.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

LOVIBOND, S. H. Sensory preconditioning: Central linkage or response mediation. *J. exp. Psychol.*, 1959, **58**, 469–475.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana, Ill.: Univer. Illinois Press, 1957.

VOLKOVA, V. D. On certain characteristics of the formation of conditioned reflexes to speech stimuli in children. *Fiziol. Zh. SSSR*, 1953, **39**, 540–548. Cited by G. Razran, The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychol. Rev.*, 1961, **68**, 81–147.

VOLKOVA, V. D. On the correction influence of the environment in elaborating conditioned reflexes to several verbal stimuli. *Zh. vyssh. nervn. Deiatel.*, 1957, **7**, 525–533. (*Psychol. Abstr.*, 32:5185)

(Received February 8, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 7–10

GEOMETRIC AND ARITHMETIC MEANS AS INDEXES OF UCS INTENSITY WITH VARIABLE REINFORCEMENT[1]

GEORGE E. PASSEY AND FRANCIS SEKYRA

University of Alabama

7 groups of 15 Ss each acquired conditioned eyelid responses to a 1,024-cps tonal stimulus 40 db. above threshold. 3 groups were given varied reinforcement with distributions generating arithmetic means of 6.5 psi and geometric means of 5.5, 4.5, and 3.6 psi, respectively. 4 control groups received constant reinforcement at 6.5, 5.5, 4.5, or 3.6 psi. 70 acquisition and 20 extinction trials were given. Each variable group performed similarly to the constant group reinforced at the geometric mean intensity of the variable group rather than the constant group reinforced at the arithmetic mean intensity of the variable group. Geometric rather than arithmetic mean intensity is the best representative stimulus value for variable reinforcement series. Groups having differences equal to or in excess of 2 psi yielded performance differences.

Variation in the magnitude of reinforcement was shown by Passey and Burns (1962) to be a determiner of acquisition rate for the conditioned eyelid response. Two variable intensity groups were compared with a constant intensity group under 100% reinforcement with the arithmetic mean intensity of the puffs constant for all groups. With increasing variability of the intensity of the reinforcing air puffs, performance in acquisition declined. Passey and Wood (1963) confirmed this finding. Arithmetic mean intensity of the air puff was held constant for each of the variable groups, though as variability was increased there was a decline in geometric mean.

Lower performance might have been a function of the increasing variability or of declining geometric mean or both since the two were confounded. The work of Helson (1959) suggests that decline in performance might accompany decline in geometric mean rather than stimulus variability per se. In a preliminary study (Passey & Sekyra, 1963) a variable reinforcement group was compared with two constant reinforcement groups, one at its arithmetic and one at its geometric mean. The curve of acquisition for the variable group was not different from that for the constant group at its geometric mean, but both of these groups differed from the constant group at the arithmetic mean. The present experiment was carried out to assess the generality of this finding.

METHOD

Seven groups of 15 *S*s each acquired conditioned eyelid responses to a 1,024-cps tone 40 db. above threshold, which served as CS. The UCS was an oxygen puff delivered to the cornea of the right eye through an aperture $\frac{5}{32}$ in. in diameter, located approximately 2 in. horizontally from *S*'s eye. The intensity of the air puff was measured at the point of delivery. A summary of the intensities of air puff used for each reinforcement group appears in Table 1. The duration of the CS was 700 msec., and the UCS was delivered during the terminal 50 msec. of CS presentation. All variable UCS schedules were presented in random sequence. The intertrial interval followed a variable sequence with a median

---

[1] This investigation was supported in part by Public Health Service Research Grant MH07390-01 from the National Institute of Mental Health and by the Research Committee of the University of Alabama.

TABLE 1

Composition of UCS Schedules for Experimental Groups

Group	Air Puff Intensity (psi)	Number of Presentations	Arith. *M* (psi)	Geom. *M* (psi)
1	6.5	70	6.5	6.5
2	1.3 5.6 6.9 8.7 10.6	15 14 14 14 13	6.5	5.5
3	5.5	70	5.5	5.5
4	.6 4.5 6.9 9.3 11.1	14 14 14 14 14	6.5	4.5
5	4.5	70	4.5	4.5
6	.6 6.9 11.1	22 22 26	6.5	3.6
7	3.6	70	3.6	3.6

intertrial interval of 19 sec., and a range of 12–26 sec. The apparatus has been described by Passey and Wood (1963).

The *S*s were recruited from introductory psychology classes and participated as part of the course requirement. Assignment of *S*s to the experimental groups was random with the restriction that sex distribution be equalized. Eight males and seven females were utilized in each group. Of the 119 *S*s recruited, 11 were rejected as sensitized responders, 1 as a voluntary responder, and 2 for apparatus failures. Criteria for rejection of sensitized and voluntary responders were those employed by Passey and Sekyra (1963).

The steps in stimulation were as follows: (*a*) Each *S* was instructed that the investigation concerned eyelid conditioning. He was asked to relax as much as possible and to make no attempt to control the movement of his eyelids. He was told that the experiment would last for about 30 min. and asked to look at an illuminated milk-glass screen directly ahead at a distance of 6 ft. for the duration of the session. (*b*) An auditory threshold was established by the method of limits and the CS intensity control was set. (*c*) The response recorder was attached to the eyelid by means of a strip of adhesive tape, the air conditioner was started, and *E* left the room. (*d*) The *S* was then stimulated; four times with CS alone; three times with UCS alone; and three times again with CS alone. (*e*) Seventy paired CS-UCS presentations and an extinction series of 20 CS presentations were given. (*f*) Upon termination of the stimulation series, *S* was interrogated so as to discover whether any extraneous cues had influenced his behavior.

Results and Discussion

The record of each *S* was analyzed by counting the number of anticipatory responses during 70 acquisition trials and the number of responses during 20 extinction trials. The acquisition data are plotted for blocks of 10 trials for the various experimental treatments in Fig. 1.

The test for differential effects of the reinforcement condition yielded an F (6, 98) = 3.24 ($p < .01$). The hypothesis of no difference in number of anticipatory responses between reinforcement conditions was rejected; therefore, differential numbers of anticipatory responses accompanied different reinforcement conditions.

The test of trend over successive trials yielded an F (69, 6762) = 12.24 ($p < .001$), and it was concluded that learning did occur.

Differences in trend produced by patterns of reinforcement yielded an F (414, 6762) = 1.02 ($p < .001$). The

TABLE 2

Value of *t* Allowing Comparisons between Pairs of Groups for Acquisition

Group	1	2	3	4	5	6
2	.87					
3	1.28	.43				
4	2.34*	1.89	1.82			
5	2.11	1.66	1.50	.00		
6	3.04**	2.76*	2.72*	1.13	1.03	
7	2.74*	2.38*	2.33*	.77	.70	.32

* $p < .05$.
** $p < .01$.

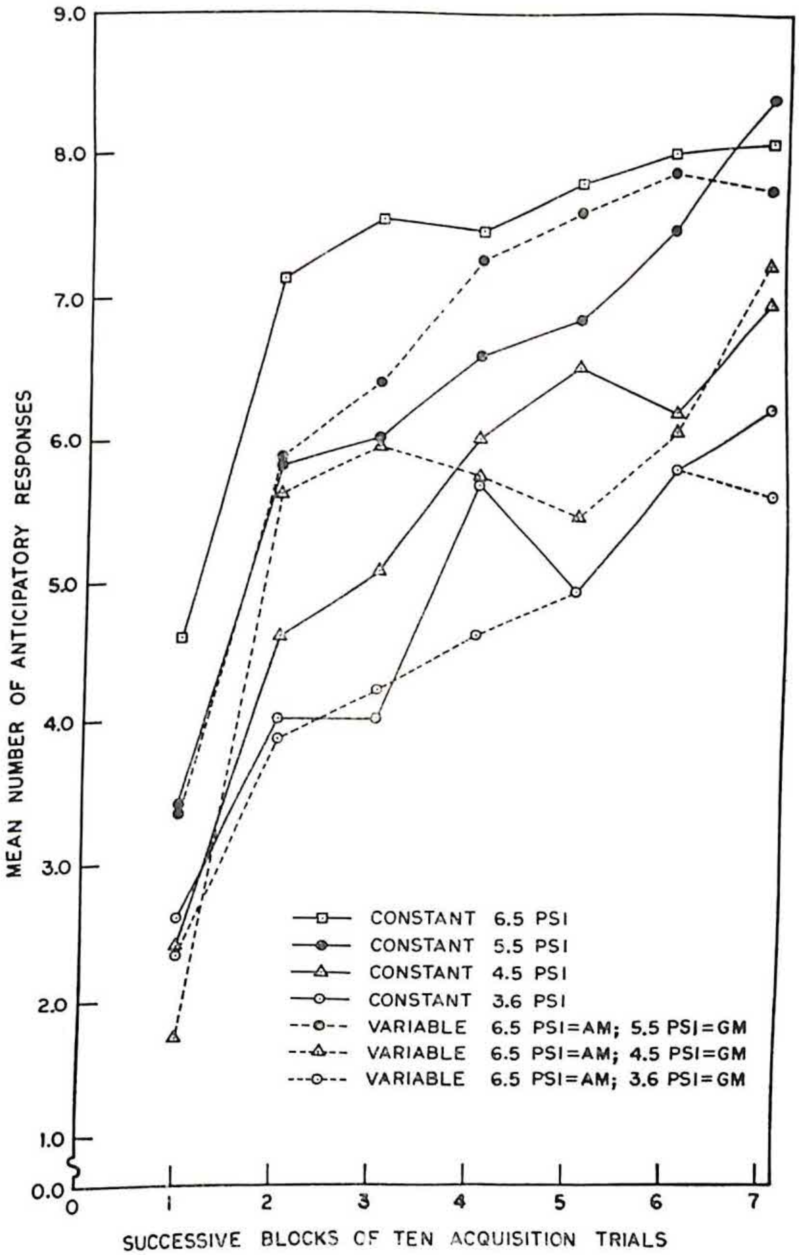


FIG. 1. Mean number of anticipatory responses for blocks of 10 trials.

hypothesis of identical trends was rejected, and it was concluded that the different patterns of reinforcement produced acquisition curves of different form.

Comparisons of each pair of experimental groups for acquisition are reported in Table 2. Comparisons between average number of anticipatory responses produced by each experimental treatment during acquisition yielded significant differences between all UCS groups where geometric mean intensities differed by

2 or more psi. A variable group tended to follow the course of acquisition of the constant group at its geometric mean intensity rather than at its arithmetic mean intensity.

Passey and Burns (1962) and Passey and Wood (1963) reported comparisons of variable reinforcement groups with constant groups at the arithmetic mean intensity for the variable groups. They attributed lower levels of performance to increasing variability. The present findings bring that interpretation into question and it appears that the effective stimulus intensity value for a variable series is its geometric rather than its arithmetic mean. This finding is in accord with the argument of Helson (1959) who holds that the best representative value of a series of stimuli is the geometric rather than the arithmetic mean.

The data for the extinction series were analyzed in the same manner as those for acquisition. An $F\ (6, 98) = 1.87\ (p > .05)$ was obtained between the experimental treatments. It was concluded that the experimental treatments failed to produce differential effects during extinction.

REFERENCES

HELSON, H. Adaption level theory. In S. Koch (Ed.), *Psychology: A study of a science*. New York: McGraw-Hill, 1959. Pp. 565–621.

PASSEY, G. E., & BURNS, T. C. The influence of variable reinforcement upon acquisition of the conditioned eyelid response. *Psychol. Rep.*, 1962, **11**, 547–552.

PASSEY, G. E., & SEKYRA, F. Variable reinforcement versus constant reinforcement in eyelid conditioning. *Psychol. Rep.*, 1963, **13**, 99–102.

PASSEY, G. E., & WOOD, D. L. Effect of pattern of reinforcement on the conditioned eyelid response. *J. exp. Psychol.*, 1963, **66**, 241–244.

(Early publication received June 18, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 11–18

HABITUATION OF THE VASOCONSTRICTIVE ORIENTING REACTION

SANFORD M. UNGER

National Institute of Mental Health, Bethesda, Maryland

Habituation of the vasoconstrictive orienting reaction was explored via an experimental paradigm in which acoustically novel stimulation characterized each successive trial: namely, numbers presented seriatim. 12 of 20 human Ss reached a habituation criterion of 3 successive nonreactions to successive numbers—followed by the immediate reinduction of vasoconstriction upon presentation of an out-of-sequence number. This phenomenon was designated selective habituation to distinguish it from the often found stimulus specificity in the habituation process. Certain of the experimental phenomena encountered, bearing on the use of photoelectric plethysmography as a research technique, were briefly noted, and some implications of selective habituation for present theorizing about habituation-mediating brain processes, with special reference to Sokolov's neuronal model, were suggested.

Habituation—that is, response decrement or cessation following repeated inconsequential stimulus presentations—is apparently a phylogenetically ubiquitous phenomenon (Hernandez-Peon & Brust-Carmona, 1961; Thorpe, 1956). In one form or another, it has been measured or observed from the sea anemone to man; in higher species, from receptor-cell activity (Livingston, 1959) to molar behavior (Hinde, 1954). Obviously, as with the more general term "learning," habituation encompasses a varied range of organismic plasticities (Razran, 1961a). While the kinds and contexts of occurrence of habituation phenomena are hardly fully charted, it is clear that nervous processes and systems differing greatly in kind and complexity are involved in their mediation (Sharpless & Jasper, 1956).

In the present study, the to-be-habituated response was finger blood vessel constriction. Finger vasoconstriction, figuring prominently in primarily Russian investigations of recent date, has been frequently used as a component indicator response for the constellation of co-occurring changes (circulatory, sense organ, skeletal muscle, EEG, cardiac, respiratory, galvanic skin, etc.) grouped together and identified as the generalized "orienting reaction" (Berlyne, 1960; Razran, 1961b; Sokolov, 1960).

In orienting reaction and other contexts, experimental habituation demonstrations have followed a relatively uniform pattern. A given stimulus display has been repetitiously presented without reinforcement, with the eventual result being a decrement or cessation of a monitored response. Subsequent stimulus presentation departing only very slightly, in any parameter, from the experimentally habituated stimulation has then been found to effect dishabituation—to be marked by the reinduction, with initial magnitude or amplitude, of whatever response. Sharpless and Jasper (1956), in underscoring this point, entitled it "stimulus specificity" in the habituation of EEG "arousal"—a finding which has been amply confirmed (John, 1961; Morrell, 1961). Studies dealing with components of the orienting reaction other than EEG arousal have also emphasized stimulus specificity and the immediate reinstatement of previously habituated orienting responses upon "intensification, weakening, or modification in any other way of any kind of stimulation [Berlyne, 1960, p. 81]."

The present procedure—representing a departure from the standard paradigm—was intended to provide an opportunity for habituation on a basis other than specificity—that is, other than after repetition of a physically identical stimulus on successive trials. Rather, each successive trial was marked by physically novel, though hardly random, stimulation. Numbers were presented seriatim: 1, 2, 3, 4, etc. In the context of such a stimulus array, presumably familiar and overlearned in *S*s' pre-experimental experiences, it was supposed that forthcoming elements in the sequence would soon be *anticipatorily* primed or triggered (see Birch & Bitterman, 1949; or Osgood's, 1957, discussion of "predictive integration"). The immediate experimental question was whether or not response decrement or cessation selectively restricted to the number *next* in sequence—that is, to the presumably expected number—would take place. Such an outcome would reflect the occurrence of habituation to stimulus presentations distinctly not specific or physically identical with earlier trials—in fact, to experimentally unique stimulation.[1]

The relevance of the present paradigm was not thought to be limited to human organisms or to be dependent on a "second signal system" level of control. In principle at least, the procedure was conceived as prototypic of any situation which contains successively different but regularly encountered or sequentially ordered, hence, familiar or expected, stimulus elements. Examples might be locomotion of an organism through a thoroughly familiar environment, or observation of the ritualized movements of an *E* preparing the apparatus prior to the one-hundredth trial.[2]

METHOD

Subjects.—Fourteen men and 6 women, ranging in age from 17–31 yr., (median, 21), served as *S*s.

Experimental setting and apparatus.—The *S*s were studied singly in a soundproof room. The *E* monitored the recording apparatus outside the room and transmitted stimuli to *S* via a microphone and earphones. Changes in finger blood volume were detected photoelectrically, with an adjustable "finger-ring" unit containing its own light source and photocell.[3] Variations in the amount of light passing through the finger, due to changes in size of blood vessels, were reflected on a thermal-pen Sanborn recorder.

Procedure.—Prior to the experimental session itself, *S*s had been informed only that the study involved physiological measures, and that electric shock would not be employed.

At the time of testing, each *S* was seated comfortably in the soundproof room with the right forearm and hand supported at about heart level. A brief explanation of the "finger-ring" plethysmograph was given to *S*, and the ring was set in place on the supported limb. The *S* was informed that the finger reaction revealed a number of physiological events, and that the purpose of the experiment was to investigate the kind of reactions elicited in a thoroughly relaxed person by sound stimuli. The task for the experiment, it was emphasized, was to remain as relaxed as possible, without moving, and with eyes closed throughout the entire session. When it was ensured that *S* fully understood the task, earphones were set in place and *E* left the room to take up his station at the recording apparatus and begin the experiment proper.

---

[1] Generalized habituation to whole classes of stimuli, for example, to words, has been reported to occur (Luria & Vinogradova, 1959). This order of consideration, not further addressed here, represents a dimension of possible complexity in habituation processes even beyond that explored via the present paradigm.

[2] Note that the careful *E*, by not altering his preparation pattern from that of previous trials, succeeds in not upsetting or alarming the "habituated" animal.

[3] This plethysmographic unit was designed and built by John A. Cooley, Technical Development Staff, Laboratory of Psychology, National Institute of Mental Health. Similar photoelectric devices are now commercially available from the Lafayette Instrument Company. For review of earlier methodology in the investigation of peripheral circulation, see Ackner (1956).

The *E* transmitted stimuli in normal conversational tone and level while continuously monitoring the blood volume record. Beginning with 1, numbers were presented seriatim at intervals ranging usually from about 5–25 sec., with the intention not to present a number when blood vessels were already, or still, constricted.

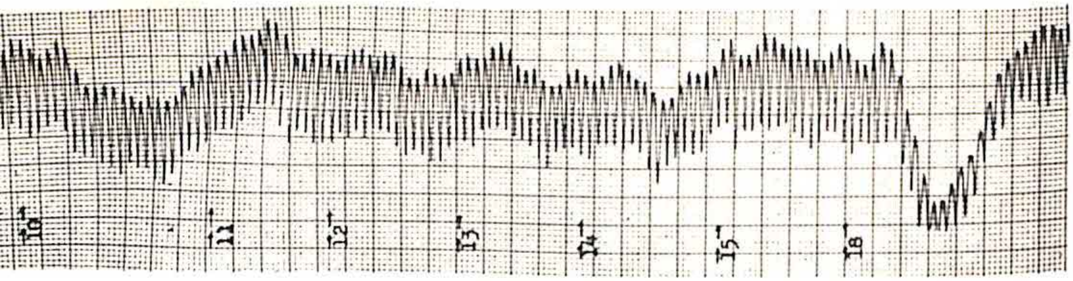
Seriatim presentations continued until the experiment was terminated, either with the conclusion that *S* was unsuitable for the purpose of the experiment or until at least three successive presentations (e.g., 14, 15, 16) *failed* to induce vasoconstrictive reactions. When this criterion of habituation was met, an out-of-sequence number was presented (e.g., 14, 15, 16, **15**; or 25, 26, 27, **29**; etc.). Subsequent to the out-of-sequence stimulus, seriatim presentations were resumed until either the experiment was terminated or until the criterion of at least three successive nonreactions was again met; at which point, another out-of-sequence number was administered. Then, the session was terminated or the procedure was maintained through additional repetitions.

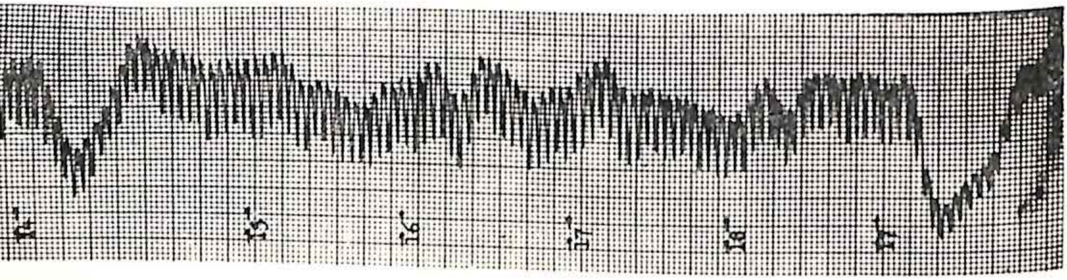
RESULTS

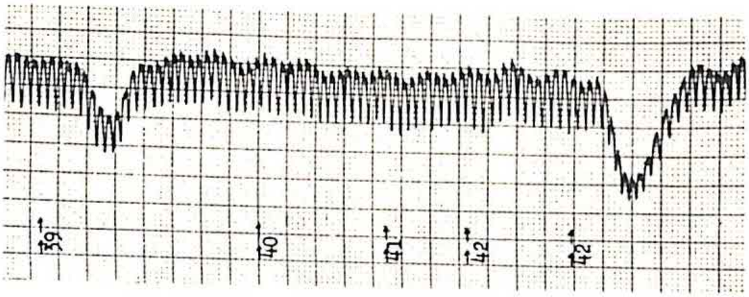
Generally speaking, both intra- and interindividual variability in vasoconstrictive responsiveness was striking, and will be briefly described below. Notwithstanding, 12 of the 20 *S*s manifested reaction patterns indicating that habituation had occurred selectively; that is, three or more successive seriatim presentations, though experimentally novel, resulted in "no reaction," followed by a clear vasoconstrictive orienting reaction upon the introduction of an out-of-sequence number. Some examples are shown in Fig. 1.

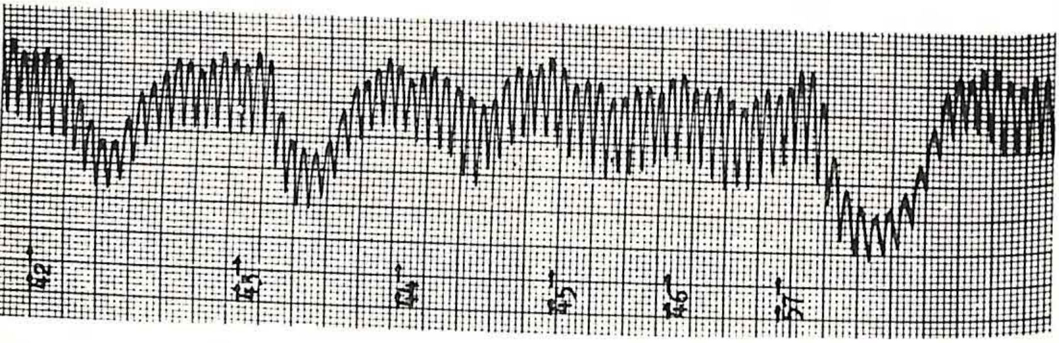
The eight *S*s in whom habituation was not demonstrated were essentially unsuitable for testing—for either of two reasons. One group (five *S*s) was vasoconstrictively unreactive throughout the course of the session; i.e., from the time of application of the first stimulus until the experiment was terminated, the record represented an unchanging, unwavering "baseline." It is not known whether these *S*s are analogous to "no alpha" EEG *S*s, or the observed nonreactivity was an artifact of the particular measurement procedure or experimental situation. A second group of three *S*s was vasoconstrictively hyperreactive, i.e., they showed no steady baseline; the finger blood vessels were apparently in a state of continuous change, quite independent of external stimulus presentations. For these latter *S*s, it was impossible to correlate applied stimulation with vasoconstrictions. Samples of record sections for these unsuitable *S*s are shown in Fig. 2.

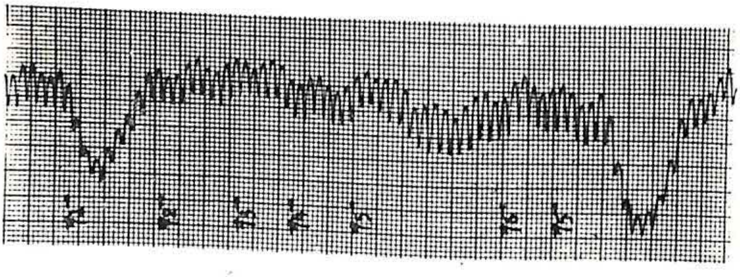
Certain other of the experimental phenomena encountered may be noted in passing, with special reference to variability in vasoconstrictive reactivity. First, for suitable *S*s, the number of trials required to reach the first habituation criterion of three successive nonreactions ranged from 9 to 36 with a median of 18. However, single nonreactions often occurred as early as the second, third, or fourth presentation. Sometimes the criterion of habituation was reached gradually, after a progressive diminution in size of constrictions; sometimes the onset was sudden, with successive nonreactions following immediately on maximal appearing constrictions. On resuming seriatim presentations subsequent to the introduction of an out-of-sequence number, reaction pat-











(Fig. 1 continued on next page.)

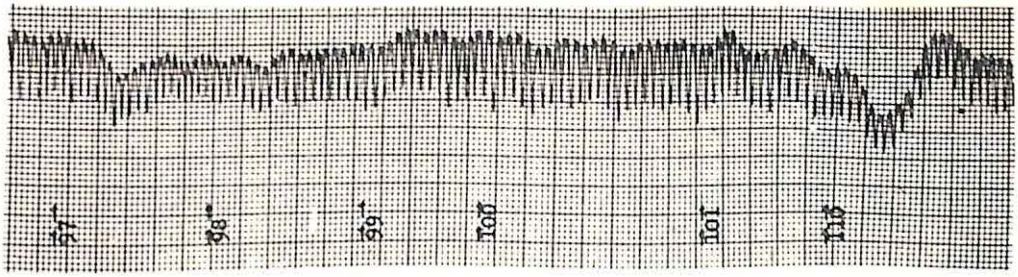
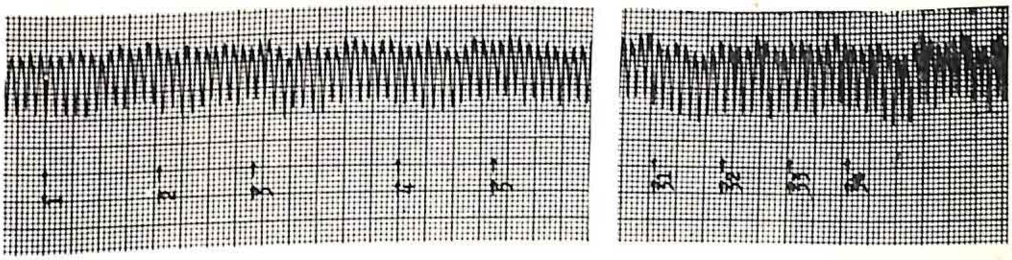
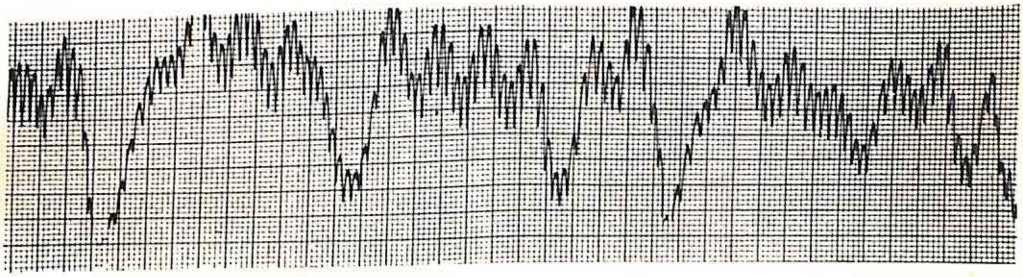


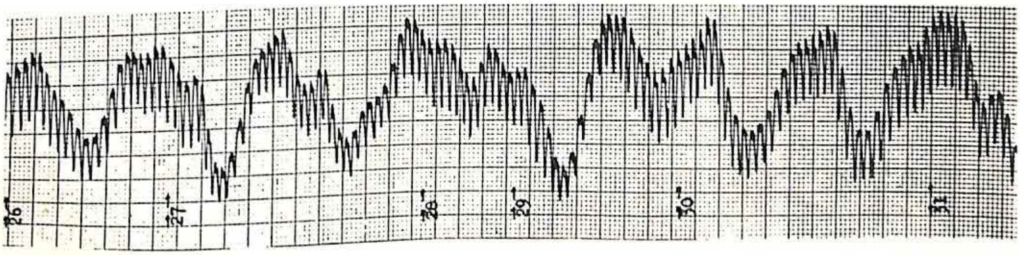
FIG. 1. Examples of selective habituation. (Paper speed for all records is 2.5 mm/sec; major vertical divisions represent 2 sec.; downward deflection indicates vasoconstriction. Each record section is from a different *S*, showing vasoconstrictive reaction to the presentation of an out-of-sequence number after at least three successive nonreactions to seriatim presentations.)



2a. Vasoconstrictively unreactive *S* at initiation of session and just prior to termination.

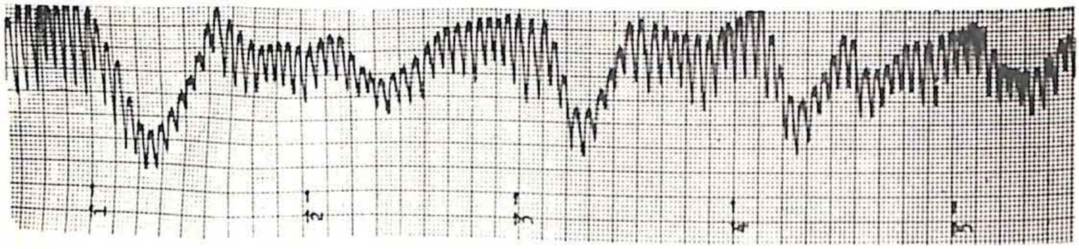


2b. Hyperreactive *S* without external stimulation.

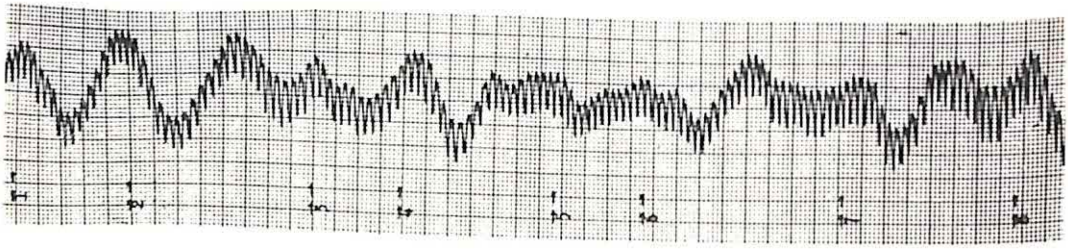


2c. Hyperreactive *S* with stimulus-uncorrelated constrictions.

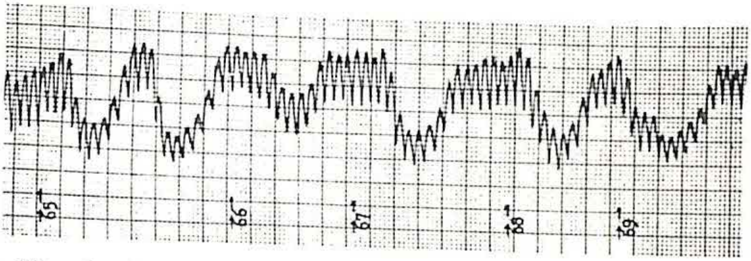
FIG. 2. Examples of "unsuitability."



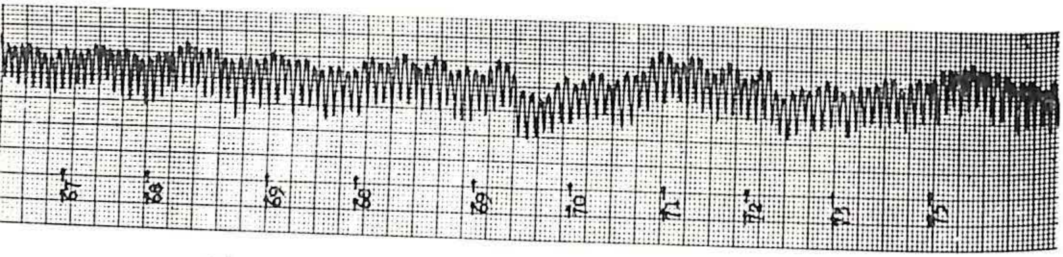
3a. Typical reactions at initiation of session.



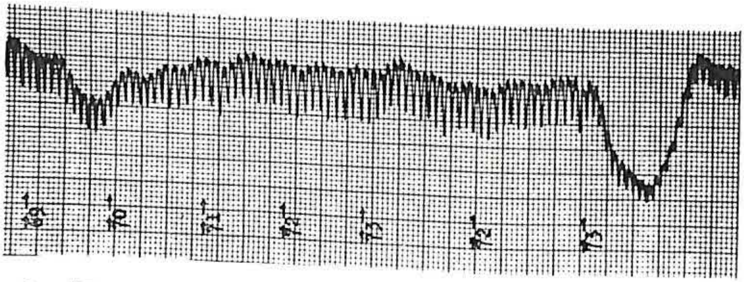
3b. Same as Fig. 3a.



3c. Nonhabituating hyperreactivity after *S* had previously habituated three times.



3d. Nonreactivity after *S* had previously been reactive.



3e. "Paradoxical" delayed reaction to out-of-sequence no. 72,

FIG. 3. Examples of variability.

terns were again highly variable. In some *S*s, selective habituation was reinstituted without difficulty, requiring fewer trials than for initial habituation, in others, more trials than initially were required; in two *S*s, three successive nonreactions never again occurred. In two others, total or generalized habituation took place; an out-of-sequence presentation no longer dishabituated an essentially unchanging nonresponsiveness. Lastly, on occasion, a paradoxical phenomenon was observed; namely, a sort of delayed dishabituation. Figure 3 contains a sampling of these events or reaction patterns.

DISCUSSION

To account for the phenomena both of habituation of the orienting reaction and dishabituation, in the experimental situations in which they heretofore have been observed, Sokolov (1960) has postulated that "a kind of neuronal model" is elaborated in the cortex, preserving information about the intensity, the quality, the duration, and other parameters of some repeated stimulus (p. 205). The development of such a neuronal model, he has further hypothesized, is attended by inhibitory corticifugal discharge to the brain stem, blocking EEG arousal and other orienting components usually triggered by collateral inputs to the reticular formation. Subsequent stimulus presentations which are accordant or coincident with the model are "filtered," and habituation is maintained; however, any discordant stimulus presentation results in dishabituating corticoreticular signals and attendant reinduction of orienting reactivity. Thus, the dishabituation of the orienting reaction is considered to be a consequence not of "novel" stimulation alone but of "impulses arising as a result of noncoincidence between a certain cortical pattern (the model) and the applied stimulation [Sokolov, 1960, p. 205]." (Cf. Magoun, 1961.)

The findings of the present study may be taken as tentatively demonstrating, at least for human organisms, that selective habituation to experimentally novel but presumably expected stimulus presentations is a phenomenon to be reckoned with. If indeed a "filtering model" of the type described by Sokolov mediates habituation, then under circumstances such as pertained in the present experiment, the "model" must be capable of successively and appropriately changing from trial to trial, of virtually reconstituting itself after each succeeding stimulus, to reflect a repeated pre-experimental experience of sequential order, of what follows what, in the stimulus world.

Needless to say, though selective habituation would appear to make good sense from a biological standpoint, the extent of its generality as a phenomenon remains an open question—one, however, that does appear eminently approachable experimentally.

REFERENCES

ACKNER, B. Emotions and the peripheral vasomotor system: A review of previous work. *J. psychosom. Res.*, 1956, **1**, 3–20.

BERLYNE, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

BIRCH, H. G., & BITTERMAN, M. E. Reinforcement and learning: The process of sensory integration. *Psychol. Rev.*, 1949, **56**, 292–308.

HERNANDEZ-PEON, R., & BRUST-CARMONA, H. Functional role of subcortical structures in habituation and conditioning. In, *Brain mechanisms and learning: A symposium.* Oxford: Blackwell, 1961.

HINDE, R. A. Changes in responsiveness to a constant stimulus. *Brit. J. anim. Behav.*, 1954, **2**, 41–45.

JOHN, E. R. High nervous functions: Brain functions and learning. *Annu. Rev. Physiol.*, 1961, **23**, 451–484.

LIVINGSTON, R. B. Central control of receptors and sensory transmission systems. In J. Field, H. W. Magoun, & V. E. Hall (Eds.), *Handbook of physiology.* Vol. 1. Washington, D. C.: American Physiological Society, 1959.

LURIA, A. R., & VINOGRADOVA, O. S. An objective investigation of the dynamics of semantic systems. *Brit. J. Psychol.*, 1959, **50**, 87–105.

Magoun, H. W. Recent contributions to the electrophysiology of learning. *Ann. N. Y. Acad. Sci.*, 1961, **92**, 818–829.

Morrell, F. Electrophysiological contributions to the neural basis of learning. *Physiol. Rev.*, 1961, **41**, 443–494.

Osgood, C. E. A behavioristic analysis of perception and language as cognitive phenomena. In, *Contemporary approaches to cognition*. Cambridge: Harvard Univer. Press, 1957.

Razran, G. Evolutionary levels of higher nervous activity. Queens College of the City University of New York, 1961. (Ditto) (a)

Razran, G. The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychol. Rev.*, 1961, **68**, 81–147. (b)

Sharpless, S. K., & Jasper, H. H. Habituation of the arousal reaction. *Brain*, 1956, **79**, 655–681.

Sokolov, E. N. Neuronal models and the orienting reflex. [In M. A. B. Brazier (Ed.), *Central nervous system and behavior*. New York: Josiah Macy, Jr. Foundation, 1960.

Thorpe, W. H. *Learning and instinct in animals*. London: Methuen, 1956.

(Received January 2, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 19–21

DOES RECODING FROM BINARY TO OCTAL IMPROVE THE PERCEPTION OF BINARY PATTERNS?[1]

E. T. KLEMMER

Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey

4 Ss were trained to report a tachistoscopically flashed binary pattern in terms of octal digits. The patterns were formed on 21 small light bulbs arranged in 7 groups of 3 each along a straight line. Little or no improvement in performance was produced by training Ss in binary-to-octal recoding even after considerable practice.

G. A. Miller (1956) in his paper on "The Magical Number Seven" cites an experiment by Sidney Smith in which the memory span for binary digits was increased from 12 to 40 by having *S* learn to recode the digits into octal and higher order codes. The present experiment seeks to determine if a similar improvement can be demonstrated in the perception of binary dot patterns.

METHOD

Apparatus.—The apparatus was an improved version of that described in a previous paper (Klemmer, 1963). Twenty-one small incandescent light bulbs were mounted on a visual perimeter 10 ft. in radius. The bulbs were arranged in seven groups of three bulbs each with ¼° between bulbs within groups and 2° between the end bulbs of adjacent groups. Thus the overall visual angle was 12½°. The center bulb was indicated by a small projecting wire. No other index marks were present. The canvas background had an apparent brightness of 1.2 ft-c which made the positions of the bulbs clearly visible before and after the tachistoscopic flash.

Stimuli.—The stimulus patterns and poststimulus cues were preprogramed on punched cards. These cards were fed into a special device which set up the patterns automatically. The stimulus patterns and sequence of patterns were completely random so that each of the eight possible three-bulb patterns (including no-bulbs-lighted) was equally likely to appear in each of the seven groups and independent of the patterns appearing in the other six groups.

Common procedure.—The *S* fixated the middle bulb of the center group and pressed a button which triggered the .1-sec. flash of the pattern. One-half second later, on those trials using a poststimulus cue, one group of three bulbs came on dimly. This group, chosen at random without regard to the stimulus pattern, cued *S* to report on the stimulus pattern which appeared on those three bulbs only. On trials in which a poststimulus cue was not used, *S* was required to report the entire pattern immediately after the stimulus flash.

The *S*'s response in all conditions was verbal. For the binary report conditions he said "one" for on and "oh" for off. For the octal coding conditions the usual translation between binary and octal numbers, 0 through 7, was used. If more than one number was required in the response, *S* was instructed to report the pattern from left to right. The responses were recorded by *E* and scored immediately by reillumination of the stimulus pattern. The *S* thus had immediate knowledge of results after each trial from seeing the stimulus pattern for several seconds.

The *S*s were normally given one session per day. Each session consisted of either 126 stimuli for the cued conditions or 48 stimuli for the full-report conditions according to the schedule given below. The sessions lasted about 40 min.

Subjects.—Four paid college students served as *S*s. All were determined to have normal vision on an orthorater test. All *S*s had previously served in an experiment on the

---

[1] The experimental trials and preliminary analysis were done at the International Business Machines Research Center. J. A. Tirrell prepared the stimulus cards and supervised the data collection. N. J. Anthony wrote the computer program. Their contributions to this study are gratefully acknowledged.

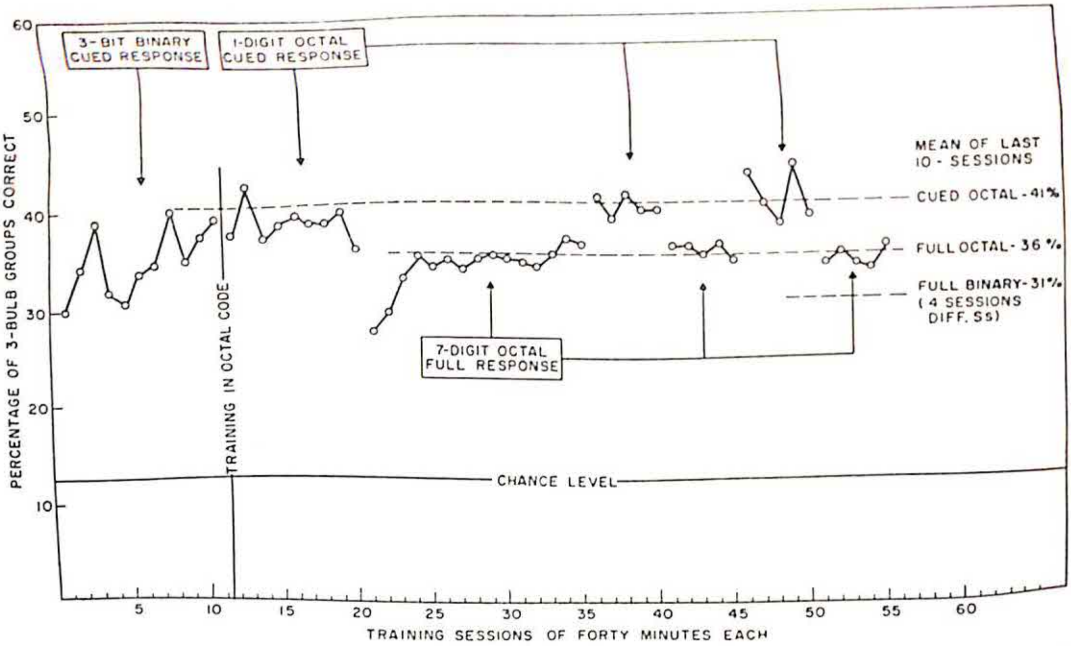


FIG. 1. Order of conditions and results in terms of percentage of three-bulb groups correct. (Each point is a mean of four *Ss* and 126 stimulus presentations per *S* for the cued conditions and 48 stimulus presentations per *S* for the full-response condition.)

same apparatus but without grouping of stimulus bulbs and with cuing by single bulbs only for a 1-bit response.

Sequence of experimental conditions.—The first sessions utilized the poststimulus cuing procedure which eliminates most of the response interference present when full report is required. The cuing technique also permits the entire response to be made a very brief time after stimulus presentation, thus keeping to a minimum any reduction in performance which is closely correlated with the passage of time. In this way the cuing method reduces or eliminates some effects which involve the memory and read-out of information and therefore is a better measure of the information initially perceived.

The general plan was to (*a*) provide enough practice with the binary reporting procedure to establish a reasonably stable level of performance; (*b*) teach recoding to octal and see if there was an acceleration or jump in performance immediately or with additional practice; (*c*) require a full response in octal digits and see if the performance difference between cued response and full report was reduced with continued training on both. This last question is relevant to the hypothesis (Klemmer, 1961) that with efficient recoding the human can transmit (i.e., put out in the response) all of the information he perceives as sampled by the poststimulus cue. The order of conditions and number of sessions under each may be read from Fig. 1. The training in binary to octal conversion occurred between Sessions 11 and 12 during a separate session using flash cards.

In addition to the above sessions, four different *Ss*, not trained in octal, were run under the requirement of full read-out with the binary response. Since these additional *Ss* were not matched in skill and training with the original *Ss*, the differences in performance must be interpreted with caution.

RESULTS

The average performance of all *Ss* on each session is given in Fig. 1 in terms of percentage of three-bulb groups correctly reported. All seven groups are equally weighted in this average since the poststimulus cue sampled uniformly over the entire field and the full-response condition sampled the entire field in each trial. Note that chance level for the 3-bit group is $(\frac{1}{2})^3$ or 12.5% correct.

From Fig. 1 it can be seen that the introduction of the octal-coded response on Session 12 produced little or no change in performance. Performance on Sessions 12–20 is a reasonable extension of the first 11 sessions. When the full 7-digit octal response was required on Session 21 performance dropped markedly but climbed up to a fairly stable value in a few sessions. Extended training on both the cued and full-response conditions through Session 55 failed to produce further improvement in either condition and thus no change occurred in the difference between them. The full report remains 5% below the cued response condition. Considering the last 10 sessions under each condition, the standard error of the mean percentage correct is considerably less than 1% for each condition based on the day-to-day fluctuations in the mean performance of the four *S*s. Thus, the difference between conditions is highly reliable.

A line is drawn on the right-hand side of Fig. 1 to indicate the average performance of the four additional *S*s who ran four sessions under the condition of full response with binary code (all 21 bits). These *S*s with considerably less training than the original four *S*s were only 5% below the full-response octal code condition.

Discussion

There is no escaping the fact that the findings of this study are negative. Extended training in binary to octal recoding did not produce measurable improvement in the perception of briefly exposed binary patterns even though the stimulus dots were conveniently grouped by threes.

It was not possible to run a control group of *S*s through an equivalent amount of training using the binary response only. Indeed, it would be questionable as to whether such a control group could be prevented from developing personal codes over the weeks of practice involved. The fact that four different *S*s with less total training, who used the binary code only, came within 5% of the performance of the octal-trained *S*s after 44 sessions of practice on the octal code, also discourages any attempt to pinpoint the difference between direct-binary and octal-recoded responses.

There is an obvious hypothesis about the negative results of the present study and the gains due to binary-to-octal recoding previously reported. It is that recoding of the type studied here takes more time than is available in the image resulting from a tachistoscopic flash. Another way of stating this is that recoding may aid the memory (storage) of information which is available long enough to be recoded but does not aid the momentary perception of such information. This, of course, is too broad a conclusion to draw from any one study.

REFERENCES

KLEMMER, E. T. The perception of all patterns produced by a seven-line matrix. *J. exp. Psychol.*, 1961, **61**, 274–282.

KLEMMER, E. T. Perception of linear dot patterns. *J. exp. Psychol.*, 1963, **65**, 468–473.

MILLER, G. A. The magical number seven, plus or minus two. *Psychol. Rev.*, 1956, **63**, 81–97.

(Received January 10, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 22–26

LIGHT INTENSITY AND BINOCULAR RIVALRY [1]

IRA T. KAPLAN AND WILLIAM METLAY

Public Service Research, Incorporated, Stamford, Connecticut

A luminous bar was presented to each eye so that the 2 bars appeared to cross in the center of the binocular field. Rivalry occurred at the intersection, where each bar seemed alternately before or behind the other. Raising the illumination of both bars increased their alternation rate to a level that held constant at higher brightnesses, until the rate declined when the light became very intense. With both bars illuminated equally, one S reported the right bar dominant more than half the time; another S reported the left bar dominant. Raising the illumination of 1 bar increased the duration that it prevailed. Prevalence increased steeply at low intensities, leveled off at higher intensities, and finally declined when the bar became glaringly bright. After viewing 2 bright bars, S reported rivalry between their afterimages, which demonstrates that rivalry can occur when the stimuli occupy fixed retinal locations.

Each eye of a person receives its own image of the scene before it. Although the two monocular views differ slightly, they coalesce into a unitary visual perception. This process is called binocular fusion. If very different patterns are presented to each eye, however, conflict results as the two views compete for attention. This condition is known as binocular, or retinal, rivalry.

Breese (1899, 1909) induced rivalry by showing a red square to one eye and a green square to the other. The *S* saw a single square that changed from red to green and back again. The colors alternated slowly in dim illumination, more rapidly when the light was bright. When one field was brighter than the other, its color was seen a greater part of time. Size and retinal location were among the other variables that influenced the results.

The present experiment re-examined the effects of intensity on binocular rivalry, using different stimuli and more light levels than Breese. A luminous bar was presented to each eye so that rivalry ensued where the bars appeared to cross: one bar seemed alternately before or behind the other. Alternation rate was measured over a range of intensities equal for both eyes. In addition, the duration that one bar predominated was determined at various intensities of that bar, while the brightness of the other bar was constant.

METHOD

The *S* looked through two 2-mm. artificial pupils at an opal glass screen 30.5 cm. from his eyes. He maintained his position with the aid of a bite bar. Mirrors and lenses focused two images of a ribbon filament (6 v., 18 amp.) on the screen. A template on the front of the screen formed the patterns shown in Fig. 1a, which were divided by a partition into two monocular fields. Wratten neutral filters in the two light paths behind the opal glass varied the illumination of each field independently. Each eye was shown a white diagonal bar between two white squares on a black ground. The left bar was tilted 45° to the left, the right bar 45° to the right. In the center of each bar was a black dot, which *S* fixated. The squares helped in achieving binocular combination of the two fields.

When the dots and squares were binocularly fused, the two bars formed an "X" with

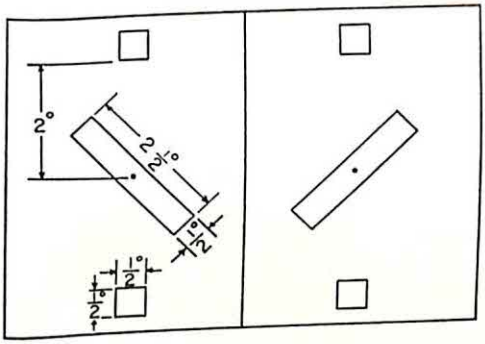
---

[1] This research was supported by a research grant (M-5364) from the National Institute of Mental Health, Public Health Service.

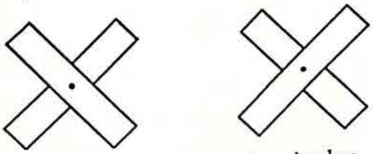
rivalry at the intersection. Only one bar appeared complete at any given moment. Usually one bar seemed in front of the other, but sometimes the recessive bar was half or wholly missing. Two stages in the rivalry cycle are illustrated in Fig. 1b: the left bar predominates in the left figure, and the right bar is dominant in the right one. When the left bar prevailed, *S* pressed a key at his left hand. He pressed another key with his right hand to indicate right-bar prevalence. A pen recorder marked the duration of each response on a moving paper tape. After adapting to each light level, *S* responded for 1 min.

RESULTS

A cycle in binocular rivalry consists of a left-dominant phase and a right-dominant phase. The *S*'s behavior can be described by two measures: number of cycles per minute and percentage of time that one field prevails over the other. Each point in the following graphs is based on the average of two 1-min. records. Light intensity is expressed as retinal illumination in $\log_{10}$ trolands. Trolands equal 2.5 × luminance in millilamberts × square of the pupil diameter in millimeters.



1a. The two monocular stimulus patterns with dimensions shown in degrees of visual angle. (The forms were white on a black ground.)



1b. Two stages in binocular rivalry. (The left bar is dominant in the left diagram; the right bar prevails in the right diagram.)

FIG. 1. Monocular stimulus patterns and binocular rivalry.

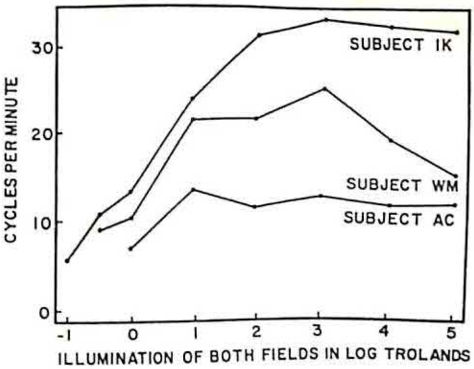


FIG. 2. Alternation rate as a function of retinal illumination, which was equal for both eyes. (Each curve was obtained from a different *S*.)

Alternation rate.—The effect of intensity upon alternation rate varied from one *S* to another. Results for three *S*s are presented in Fig. 2. Cycles per minute is plotted against log trolands, which was equal for both monocular fields. The curves begin at the lowest intensity where fusion could be maintained, which was different for each *S*. IK's rate showed the greatest increase, rising from 7 cpm at −1 log troland to a maximum of 33 cpm at 3 log trolands. Most of the increase occurred below 2 log trolands. AC's rate changed least. It rose from 7 to 13 cpm over the intensity range from 0 to 1 log troland, and became no faster at higher brightnesses. WM's curve shows a moderate rise followed by a distinct decline above 3 log trolands.

After looking at two intense fields, *S* saw rivalry between the monocular afterimages. The afterimage bars alternated where they crossed, like the luminous bars but more slowly. Subsequent to viewing both fields at 5 log trolands, IK reported after-

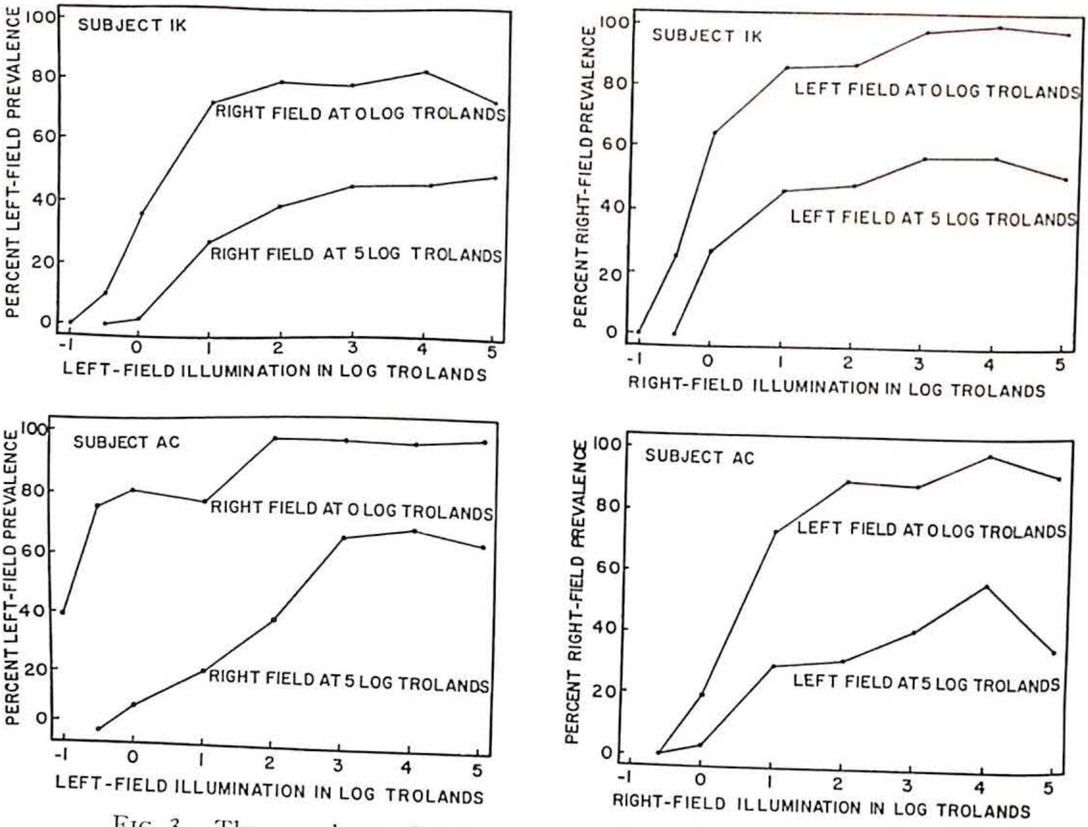


FIG. 3. The prevalence of each monocular field as a function of its intensity with the other field constant at two different levels. (Data from two *Ss.*)

image rivalry at 7 cpm, WM reported 6 cpm, and AC reported 3 cpm.

When the illumination of only one field was varied, reducing intensity tended to lower alternation rate, and increasing intensity generally raised alternation rate until the brighter field became prevalent most of the time and reversals grew fewer.

Prevalence

Raising the illumination of either field increased the duration that it dominated. The upper right graph in Fig. 3 shows, for IK, how the prevalence of the right field increased as its intensity was raised from −1 to 5 log trolands—from very dim to glaringly bright. The two curves were obtained with the left field constant at two different levels. When the left field was kept at 0 log trolands, the prevalence of the right field reached 100%. With the left field at 5 log trolands, right-field prevalence did not exceed 57%. Both curves rise to plateaus and decline slightly at the highest intensity.

Alternation occurred even when the monocular fields were very different in brightness. For instance, when IK's right field was only 0 log trolands, it prevailed 27% of the time over the left field at 5 log trolands. The visual appearance of a dim bar even briefly predominating over a much brighter one was rather striking.

IK's right eye tended to dominate his left. When both monocular fields were equally bright, the right field prevailed more than half the time. Over a 6 log unit range of equal intensities, right-field prevalence averaged 59%. The dominance of IK's right eye is shown also by comparison of the two upper graphs in Fig. 3. Variables that are plotted for the right field on one graph are plotted for the left field on the other. Right-eye dominance is indicated by the fact that the curves of right-field prevalence are higher than the corresponding curves of left-field prevalence.

AC was left-eye dominant: his left-prevalence curves in the lower left graph in Fig. 3 are generally higher than his right-prevalence curves in the adjacent graph. The upper left-prevalence curve could not be reduced to 0% because AC was unable to maintain fusion when the left field was dimmer than −1 log trolands. Data obtained with both fields equally bright at various levels between 0 and 5 log trolands, showed the left field prevailing 64% of the time. AC's curves resemble those of IK in that their initially positive slopes become level and then usually reverse at the highest intensities.

Discussion

This study examined the influence of light intensity on binocular rivalry. Alternation rate was measured with both monocular fields equally bright, and the prevalence of one field was determined as a function of its intensity while the other field was held constant. Rivalry between afterimages was incidentally observed. The results confirm Breese's observations, but in addition they show trends not apparent in his data, trends that appear also in the effects of illumination on stereoscopic acuity, visual acuity, and critical flicker frequency.

Breese found that alternation became faster when both fields were made brighter, and that a brighter field prevailed over a dimmer one a major part of the time. The present experiment defined the effects of illumination more fully by testing more intensity levels. Thus, it was seen that the rise in alternation rate and in prevalence occurred primarily below 1 or 2 log trolands. Further increments in brightness produced little additional rise, and above 4 log trolands often produced a decline. These results were for cone vision, since fixation was central and intensities were in the photopic range.

Stereoscopic acuity also rises to an asymptotic level. Mueller and Lloyd (1948) measured the effect of increasing illumination on threshold binocular disparity, which is the inverse of stereoscopic acuity, i.e., lower disparity signifies higher acuity. The cone branch of the disparity curve declined to a level that was maintained over the upper range of photopic illumination. Very high intensities, which might have produced a final increase in disparity thresholds, were not used.

Illumination has a similar effect on visual acuity, although the exact form of the curve varies with the test object. Shlaer (1937) obtained slightly different results with a grating than with a C, but in both cases when acuity was plotted against log trolands, the cone-vision curves rose steeply at first and then became nearly horizontal. Under certain conditions, increasing intensity reduces acuity after raising it to a maximum (Fry & Cobb, 1935; Lythgoe, 1932; Wilcox, 1932).

Critical flicker frequency lacks a broad level range, but does decline after reaching a maximum. Hecht (1938) presented data obtained with a flickering field that stimulated the foveal cones. The resultant curve of critical flicker frequency vs. log retinal illumination rose linearly to a peak at about 4 log trolands and then turned down.

The occurrence of rivalry between afterimages is interesting in relation to the role of eye movement in binocular rivalry. A hypothesis that rivalry changes are caused by the shifting of retinal images to different receptors was contradicted by Peckham's (1936) observation of only chance coincidence between alternations and eye movements. Afterimage rivalry also is inconsistent with this hypothesis, since afterimages occupy fixed retinal locations.

REFERENCES

BREESE, B. B. On inhibition. *Psychol. Rev. mongr. Supple.*, 1899, **3**(1, Whole No. 11).

BREESE, B. B. Binocular rivalry. *Psychol. Rev.*, 1909, **16**, 410–415.

FRY, G. A., & COBB, P. W. A new method of determining the blurredness of the retinal image. *Trans. Amer. Acad. Ophthalmol. Otolaryngol.*, 1935, **40**, 423–428.

HECHT, S. The nature of the visual process. *Bull. N. Y. Acad. Med.*, 1938, **14**, 21–45.

LYTHGOE, R. J. The measurement of visual acuity. *Med. Res. Council spec. rep., Ser.*, 1932, No. 173.

MUELLER, C. G., & LLOYD, V. V. Stereoscopic acuity for various levels of illumination. *Proc. Nat. Acad. Sci., Washington*, 1948, **34**, 223–227.

PECKHAM, R. H. Eye movements during "retinal rivalry." *Amer. J. Psychol.*, 1936, **48**, 43–63.

SHLAER, S. The relation between visual acuity and illumination. *J. gen. Physiol.*, 1937, **21**, 165–188.

WILCOX, W. W. The basis of the dependence of visual acuity on illumination. *Proc. Nat. Acad. Sci., Washington*, 1932, **18**, 47–56.

(Received January 22, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 27–32

SOME EFFECTS OF MONETARY REWARD AND KNOWLEDGE OF RESULTS ON JUDGMENT[1]

GEORGE S. LARIMER AND B. JACK WHITE

University of Utah

3 groups judged a series of weights in a treatment and an anchoring session. In the former, 1 group served with no knowledge of results (NKOR), 1 received KOR, and 1 received monetary rewards (MR). Accuracy was greater with KOR or MR than with NKOR; however, effects of MR were not greater than KOR. In the anchoring session, MR and KOR were discontinued and the original series was judged in the presence of a remote anchor. No change in judgments of the series occurred in the MR group, whereas the other groups exhibited contrast. Thus, value appeared to produce resistance to change. No differences in judgments of the anchor were found.

The present study compared the effects of no knowledge of results (NKOR), knowledge of results (KOR), and monetary rewards (MR) on the formation and change of judgmental scales for a series of weights. Evidence from studies comparing the effects of KOR and MR is contradictory—some studies indicate that MR produces performance inferior to that produced by KOR (Miller & Estes, 1961) or, at best, equal to that produced by KOR (Estes, Miller, & Curtin, 1962), while others indicate that MR produces greater learning (e.g., Siegel & Goldstein, 1959). The bulk of evidence and theory indicates that differential amounts of reinforcement should produce differential learning unless, as Miller and Estes (1961) point out, the incentive becomes large enough to render *S* anxious or distract him. It follows that in the Formation phase of the present experiment, the greatest increase in accuracy should occur when *S* is given a moderate MR for correct judgments (e.g., a 5¢ reward such as that used by Siegel & Goldstein, 1959), that an intermediate increase should occur when *S* is given only KOR, and that the least increase should occur when *S* receives NKOR.

The Change phase of the experiment was designed to test some common assumptions concerning the effects of value on resistance to change. Persons doing attitude research generally assume that repeated rewards for "correct" judgments of social stimuli lend value to the individual's judgmental standards which, in turn, produces resistance to changing the standards. Attitude change research supports such reasoning—the greater the subjective value associated with a judgmental standard, the greater the resistance to change when propaganda which is discrepant from the individual's standard is presented (Sherif & Hovland, 1961). If the analogy holds, it follows that *S*s in the present experiment who previously received MR for correct judgments should be more resistant to change when a remote anchor is presented than *S*s who received only KOR (which, of course, may also be rewarding but presumably to a lesser degree) and they, in turn, should be more resistant to change than *S*s who received NKOR.

---

[1] This report is based on a Master of Arts thesis submitted to the faculty at the University of Utah by the first author. Financial assistance for the study was provided by the University of Utah Faculty Research Fund in a grant (No. 53-0170) to the second author.

Judged differences between valued and neutral stimuli tend to be larger than judged differences between valued stimuli or between neutral stimuli (Dukes & Bevan, 1952; Tajfel, 1959; Tajfel & Cawasjee, 1959). Thus, it would appear that in the Change phase of the experiment judged differences between a nonvalued remote anchor and the nearest stimulus in the series should be greatest for *S*s who previously received MR for correct judgments of the series, should be intermediate for *S*s who received KOR, and should be least for *S*s who received NKOR.

METHOD

Subjects.—Several criteria, discussed later, were employed to select and assign *S*s to one of three groups of 11 *S*s each. The 33 *S*s were right-handed male and female volunteers from introductory psychology classes who received points toward their course grade for participating in the study.

Stimuli.—In order to insure the occurrence of contrast effects, part of a stimulus series and an anchor which previously had been demonstrated to produce contrast (Rogers, 1941; Sherif, Taub, & Hovland, 1958) were used. The series consisted of weights of 75, 93, 109, 125, and 141 gm. (hereafter referred to as 8, 9, 10, 11, and 12, respectively). A 288-gm. weight was used as a remote anchor.

Apparatus.—Since the apparatus is described in detail elsewhere,[2] only a brief account of its functions will be given. The basic unit of the apparatus consists of a large motor-driven platform which revolves within a seven-sided enclosure. As many as seven *S*s who are seated around the sides of the apparatus put their hands into the enclosure and lift one of 40 weights which are arranged along the edge of the platform. Except for the instructions which are given to *S*s, the entire procedure is automated; the apparatus times the stopping and starting of the rotation of the platform, initiates and terminates a 1-min. rest interval after each series of 25 judgments, records *S*'s response to each stimulus as he depresses a button on his response panel, and operates a slave unit which automatically dispenses MR for each correct judgment, or turns on a red light for KOR, depending upon the experimental condition. In addition to providing machine pacing during the training session, the apparatus allowed self-pacing during the treatment and anchoring sessions.

Procedure in the training session.—In the training session, *S*s were machine paced at a rate of 8.5 sec. between stimulus presentations. The *S*s were told, in brief, that they were to judge five weights, in ounces, by pressing the appropriately numbered button (8–12) on their response panels. They were then shown the operation of the apparatus, were told to lift with a uniform wrist motion, and were given 25 practice trials. Following the practice trials, 100 judgments were obtained. Any *S* who was unable to make at least 35 correct judgments in the series of 100, χ^2 (4) = 11.25, $p < .025$, was not used in the remainder of the experiment. Of the 111 *S*s who participated in the training session, 18 failed to meet the criterion.

The 93 *S*s who did meet criterion were used to form three matched groups of 11 *S*s each. Matching on the basis of training session rate of learning was performed by the following method. Accuracy scores were obtained for each *S* for each of the four blocks of 25 trials in the training session. The four scores for each *S* were matched with those of two other *S*s in such a manner that none of the four scores per *S* differed by more than three correct judgments from comparable scores for the other two *S*s. The *S*s in each matched set of three were then randomly assigned to three experimental groups which were used in the treatment and anchoring sessions.

Matching was necessary to control for differential learning rates and discrimination abilities. Had these variables not been controlled, (*a*) differential learning rates in the three groups during the treatment session might be the result of one group having had a lower or a higher rate of learning at the outset, and (*b*) differential resistance to change during the anchoring session might be explained by one of the groups having had a higher level of discrimination at the outset.

---

[2] A 17-page report giving details concerning construction of the apparatus has been deposited with the American Documentation Institute. Order Document No. 7439 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.75 for microfilm or $2.50 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

Procedure in the treatment session.—The *S*s were individually exposed to the treatment session approximately 1 wk. after completion of the training session. Again, 100 judgments were obtained. In all groups, *S*s were told to continue using 8–12 oz. for their judgments. The NKOR group was given no further instructions. Each *S* in the KOR group was told, in addition, that the machine beside him would come on when he made a correct judgment, that he would hear it operating, and that the light on top of the machine would come on. The MR group was further informed that when the machine came on it would discharge 5 pennies as a reward for making a correct judgment, and that it was therefore possible to earn $5.00 during the next half hour.

Procedure in the anchoring session.—Following the treatment session, *S* left the room for a 20-min. rest. During this time *E* interpolated 15 anchor weights among 25 weights from the original series. At the beginning of the anchoring session all *S*s were told that, "You may, if you wish, use any number from 8–19 to describe the weights." Each *S* then made 100 judgments of the original series and 60 judgments of the remote anchor. The *S*s did not receive MR or KOR during the session.

RESULTS

Comparison of training session judgments.—Since differences between groups in training session performance could be crucial to analysis of the treatment session data, a simple analysis of variance was performed to determine whether the criteria used for matching *S*s had, in fact, been sufficiently stringent to insure statistically matched groups. The results showed that the mean accuracy scores of the groups were not significantly different from each other, $F\ (3, 40) = .052,\ p > .05$.

Comparison of increases in accuracy in the treatment session.—It was expected that the MR group would evidence a greater increase in accuracy during the treatment session than the KOR group and that the NKOR group would show the least increase. In order to test this expectation, *S*'s training session accuracy score was subtracted from his treatment session accuracy score and the resulting accuracy difference scores of the three groups were compared with a simple analysis of variance. Results indicated that the means of the three groups were not equal, $F\ (2, 30) = 3.95,\ p < .05$. An individual degrees of freedom test indicated that the NKOR group showed less increase in accuracy than the averaged increase in the KOR and MR groups, $F\ (1, 30) = 7.79,\ p < .01$. The KOR and MR groups did not differ from each other, $F\ (1, 30) = .106,\ p > .05$.

Comparison of differences between treatment and anchoring session judgments.—Mean judgments in the treatment and anchoring sessions are portrayed in Fig. 1. In order to compare statistically the amount of change produced by the anchor in each group, a change score was obtained for each *S* by computing his mean judgment of the original series in the treatment session, his mean judgment of the original series in the anchoring session, and subtracting the former from the latter. Thus, a positive change score reflects a shift in judgments toward the stimulus value of the remote anchor, while a negative score, of course, reflects the reverse.

The significance of the change scores was separately assessed for each group in order to determine whether the anchor actually produced contrast in any of the groups. As Table 1 shows, *t* tests for related groups indicated that the NKOR and KOR groups shifted their judgments away from the anchor (contrast). Despite the fact that the MR group appears to have shifted their judgments to-

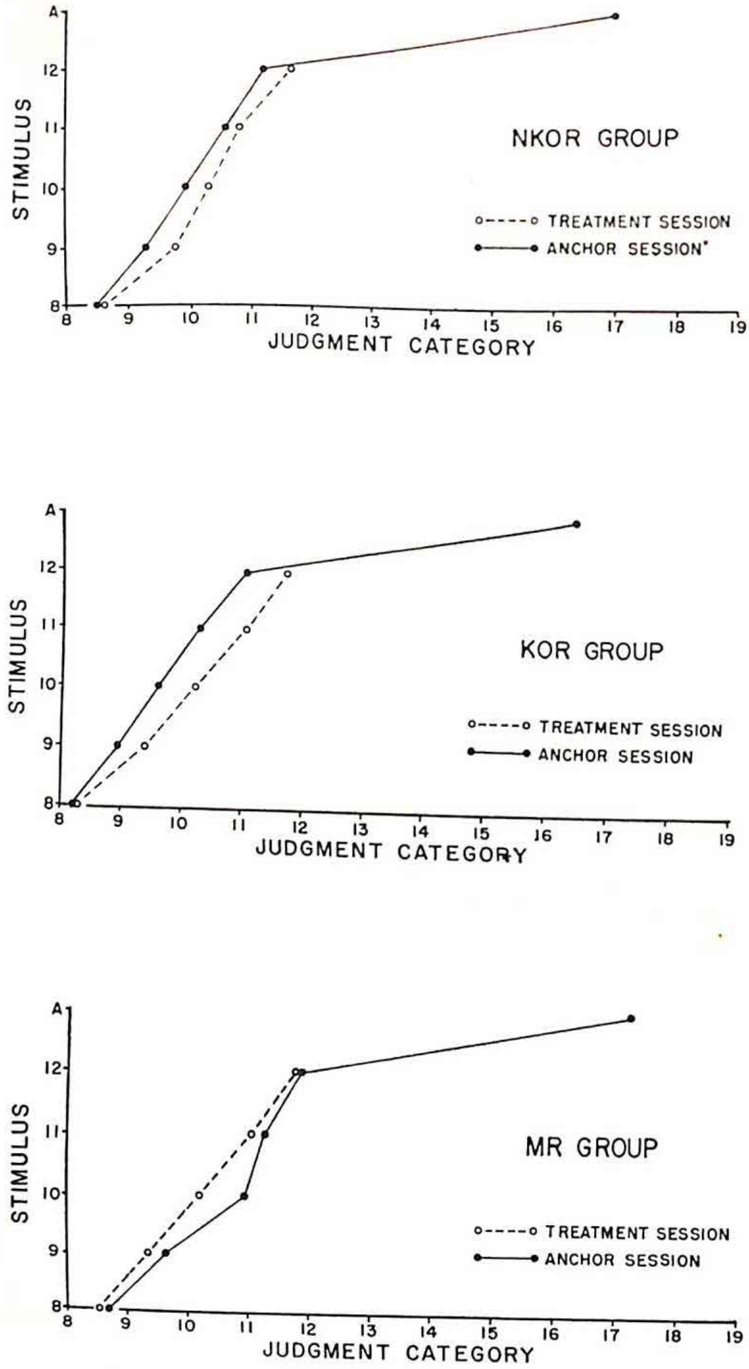


FIG. 1. Mean judgments of NKOR, KOR, and MR groups. (The anchor is denoted by A on the ordinate.)

ward the anchor in Fig. 1, the change was not significant.

The relative amounts of change in the three groups were compared by a simple analysis of variance which indicated that the means of the groups were not equal, $F\ (2, 30) = 5.82$, $p < .01$. An individual degrees of freedom test indicated that the MR group exhibited significantly less

TABLE 1

Comparison of Treatment and Anchoring Session Judgments in the Three Groups

Group	Change Score	t	df	p[a]
NKOR	−.33	2.29	10	<.025
KOR	−.50	5.19	10	<.0005
MR	.17	.93	10	>.15

[a] One-tailed probability values.

change than the averaged change in the other groups, $F\ (1, 30) = 10.91$, $p < .005$. Change scores for the NKOR and KOR groups were not significantly different, $F\ (1, 30) = .698$, $p > .05$.

Analysis of judgments of the anchor.—It was expected that the greater the amount of value associated with the stimulus series, the greater the magnitude of the difference between judgments of Weight 12 and the anchor. In order to test the expectation, a discrepancy score was obtained for each *S* by subtracting his mean judgment of Weight 12 from his mean judgment of the anchor. A simple analysis of variance performed on the discrepancy scores for the three groups indicated no overall difference, $F\ (2, 30) = .233$, $p > .05$. An individual degrees of freedom test indicated that anchor displacement in the NKOR group was not different from the averaged displacement in the KOR and MR groups, $F\ (1, 30) = .200$, $p > .05$. Further, the KOR and MR groups were not different from each other, $F\ (1, 30) = .265$, $p > .05$.

Discussion

During the Formation phase, MR did not produce a greater increase in accuracy than KOR. As indicated previously, Siegel and Goldstein (1959), using *S*s similar to those in the present study (college students in both cases), found that 5¢ rewards in a two-choice guessing game produced greater accuracy than KOR. The discrepancy in the results of the two studies may be attributable to differences in the task or in the manner in which rewards were dispensed. In the present study, a machine automatically dispensed coins to *S*s; in Siegel and Goldstein's study *E* handed coins to *S* and, thereby, may have associated social rewards with correct responses. Miller and Estes (1961) and Estes, Miller, and Curtin (1962), who also failed to find the expected differences between the effects of KOR and MR on learning, also used a machine to dispense coins. In any event, the present study indicates that the addition of MR to KOR may be an ineffective technique of increasing accuracy during the learning of a judgmental scale. This, however, is not to say that MR had no effect on learning; MR did produce effects during the Formation phase which were not evident until the Change phase.

In the Change phase of the experiment, the NKOR and KOR groups exhibited the contrast effects typically produced by remote anchors. The fact that the two groups did not differ in amount of contrast indicates that in the presence of remote anchors, subjective judgmental scales formed with KOR function in a manner similar to those which are formed with NKOR. The MR group, on the other hand, exhibited no significant change in mean judgments during the anchoring session.

Obviously, giving *S* 5¢ for a correct judgment is not identical to a family or peer group giving social rewards for "correct" judgments. The lack of change in the MR group does, however, lend tangential support to the earlier mentioned assumption concerning attitude formation; namely, that one of the factors which makes value-laden judgmental standards resistant to change is the recurrent association of rewards with appropriate judgments.

A definitive statement cannot be made regarding the effect of value on the definition of a remote anchor. All three groups estimated the anchor to be the same relative distance from the heaviest weight in the original series. In studies reporting the functioning of attitudes in situations analogous to the one used in the present experiment (e.g., Hovland, Harvey, & Sherif, 1957; Hovland & Sherif, 1952) remote anchors were judged as more remote by *S*s who held intense attitudes than by *S*s who held less intense attitudes. If the analogy is appropriate, an insufficient amount of value may have been associated with the judgmental scale in the MR group. The lack of differential anchor displacement in the three groups might also be due to the fact that the anchor was not negatively valued—the previously mentioned attitude studies used remote anchors which were negatively valued by *S*s. On the other hand, it may be that value has nothing to do with anchor displacement; the crucial variable may be amount of practice with the judgmental scale. Harvey and Caldwell (1959), Harvey and Campbell (1963), and Harvey and Rutherford (1958) found that "strong concepts" which were produced through extra practice produced more displacement in judgments of remote anchors than did "weak concepts."

REFERENCES

DUKES, W. F., & BEVAN, W., JR. Accentuation and response variability in the perception of personally relevant objects. *J. Pers.*, 1952, **20**, 457–465.

ESTES, B. W., MILLER, L. B., & CURTIN, M. E. Supplementary report: Monetary incentive and motivation in discrimination learning: Sex differences. *J. exp. Psychol.*, 1962, **63**, 320.

HARVEY, O. J., & CALDWELL, D. F. Assimilation and contrast phenomena in response to environmental variation. *J. Pers.*, 1959, **27**, 125–135.

HARVEY, O. J., & CAMPBELL, D. T. Judgment of weight as affected by adaptation range, adaptation duration, magnitude of unlabeled anchor, and judgmental language. *J. exp. Psychol.*, 1963, **65**, 12–21.

HARVEY, O. J., & RUTHERFORD, J. Gradual and absolute approaches to attitude change. *Sociometry*, 1958, **21**, 61–68.

HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 244–252.

HOVLAND, C. I., & SHERIF, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47**, 822–832.

MILLER, L. B., & ESTES, B. W. Monetary reward and motivation in discrimination learning. *J. exp. Psychol.*, 1961, **61**, 501–504.

ROGERS, S. The anchoring of absolute judgments. *Arch. Psychol., N. Y.*, 1941, No. 261.

SHERIF, M., & HOVLAND, C. I. *Social judgment.* New Haven: Yale Univer. Press, 1961.

SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55**, 150–155.

SIEGEL S., & GOLDSTEIN, D. A. Decision-making in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, **57**, 37–42.

TAJFEL, H. The anchoring effects of value in a scale of judgments. *Brit. J. Psychol.*, 1959, **50**, 294–304.

TAJFEL, H., & CAWASJEE, S. D. Value and the accentuation of judged differences: A confirmation. *J. abnorm. soc. Psychol.*, 1959, **59**, 436–439.

(Received January 22, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 33–40

EXTERNAL INHIBITION OF THE CONDITIONED EYELID REFLEX[1]

H. S. PENNYPACKER[2]

Duke University

Presentation of an extraneous auditory stimulus at the midpoint of the CS-UCS interval is shown to inhibit the blink reflex conditioned to a light. Although such inhibition appears to be unrelated to the number of prior reinforcements, its occurrence in individual Ss is inversely related to the total number of CRs produced. Evidence is also presented to suggest that the same stimulus which acts as an external inhibitor in acquisition functions as a disinhibitor if presented following partial extinction, but this suggestion is attenuated by the absence of proper controls. The results are discussed in terms of their providing a technique for examining the Pavlovian implications of these phenomena at the human level.

Among the numerous phenomena identified and investigated by Pavlov, one of the most pervasive was external inhibition: the process whereby presentation of a novel or unusual stimulus disrupts or suppresses ongoing reflex activity. Thus, if a state of excitation exists with respect to a certain reflex, application of an external inhibitor would cause the reflex to occur with reduced strength or perhaps not to occur at all. If, on the other hand, a state of inhibition exists with respect to the reflex, presentation of the external inhibitor would, according to Pavlov, inhibit the inhibition and the reflex would occur with increased vigor, giving rise to the phenomenon of disinhibition. From this it follows that the same novel stimulus can on one occasion in an experiment inhibit a conditioned reflex though it may facilitate the occurrence of the reflex on another occasion. Thus, Pavlov (1927) writes

> We have seen that the very same extra stimuli which when they evoke strong, extraneous reflexes produce external inhibition of the positive conditioned reflexes, produce, when the effect is weak from the start or weakened by repetition, disinhibition of the conditioned reflexes which were made to undergo extinction [p. 67].

These relationships have received surprisingly little systematic attention from American investigators. In 1936, M. A. Wenger conditioned the GSRs of human *S*s to a red light using a shock to the sole of the right foot as reinforcement. Twice during acquisition, he presented tactual vibration to *S*'s left hand. He then recorded the amplitude of the GSR made to the next presentation of the light CS and found in each case that it was smaller than the response made to a nonreinforced CS presented immediately prior to the vibrator. After extinguishing the response to its level prior to reinforcement, he again administered the tactual vibration and observed an increment in the GSR made to the immediately succeeding presentation of the CS. More recently, it has been noted that presentation of a "ready signal" will depress performance in an eyelid conditioning situation (Dufort & Kimble, 1958; McAllister & McAllister, 1960). All of these experiments, however, differ from Pavlov's in one important procedural detail; Pavlov presented his external inhibitors during the operation of the CS and therefore, according to him, at a time when excitatory or inhibitory states under the control of the CS were in effect. The other investigators mentioned applied external inhibitors at times when the CS was not present though it is possible, particularly in the case of the Wenger experiment, to assert that lingering excitatory and inhibitory effects developed in the presence of the CS were effected by the extra stimulus.

---

[1] Adapted from a dissertation submitted to the Department of Psychology of Duke University in partial fulfillment of the requirements of the PhD degree. The writer is greatly indebted to Gregory A. Kimble for his guidance and encouragement in the conduct of this research. This research was supported in part by United States Public Health Service Predoctoral Fellowship No. MPM-13, 327-C1 from the National Institute of Mental Health and in part by National Science Foundation Grant No. NSF-6-7079 to Gregory A. Kimble.

A briefer report of this research was presented before the Psychonomic Society, St. Louis, September 1962.

[2] Now at the University of Florida.

A few American investigators have actually presented external inhibitors during the operation of the CS (Rodnick, 1937; Spence & Runquist, 1958) but these investigations have invariably been concerned with the nature of the response made to the external inhibitor, usually a very weak stimulus used as a probe, rather than with the effects of the extra stimulus upon the primary reflex.

Finally in a preliminary study in this laboratory, it was found that a novel visual stimulus presented instead of the usual visual CS, after 15, 30, and 60 reinforcements, inhibited the conditioned eyeblink and that the degree of inhibition varied inversely with the number of prior reinforcements. In contrast to Wenger's findings, however, was the suggestion that the presentation of the novel stimulus during acquisition was followed by a period of heightened excitation, a phenomenon labeled induction by Pavlov. Unfortunately, all *Ss* in this experiment were administered all three trials involving the novel stimuli; hence, no controls were available against which to assess the presence or amount of such inductive effect. The present study was conducted in an effort to correct this failing as well as to verify the inhibiting properties of a novel stimulus presented *with*, rather than instead of, the conditioned stimulus. Since the primary interest was in an inhibitory rather than a disinhibitory effect, it was decided that the extra stimulus should be presented during the second, usually excitatory, phase of the interval of delay rather than throughout the interval as Pavlov had typically done. This procedure was also adopted with the expectation that as a result of inhibiting the excitatory phase in this manner, inhibition of the UCR might also be observed (Kimble & Ost, 1961; Kimmel & Pennypacker, 1962).

Finally, to substantiate Pavlov's and Wenger's assertion that the same extra stimulus which serves as an inhibitor during acquisition will, in the presence of inhibition acquired as a result of experimental extinction, act as a disinhibitor, a series of post-extinction trials was administered with the extra stimulus added as it had been during acquisition.

METHOD

Apparatus.—The *S* was comfortably seated alone in a well-lighted, sound-deadened cubicle whose interior dimensions were 84 × 36 × 84 in. Immediately in front of *S* was a tablelike shelf upon which he could rest his hands and arms. Communication between *S* and *E* was available when needed through an intercommunication system.

All responses were recorded by means of a microtorque potentiometer attached mechanically to *S*'s right eyelid; movements of the potentiometer arm were relayed to a Grass Model 5E dc driver amplifier through a 22.5-v. balanced connecting circuit designed and supplied by the Grass Instrument Company. The output of the driver amplifier was ink recorded on one channel of a Grass Model 5 polygraph at a paper speed of 30 mm/sec; additional pens marked the onset and offset of the CS and UCS.

The unconditioned stimulus, a puff of dry compressed air, was delivered at an intensity equivalent to 180 mm. of mercury through a $\frac{3}{64}$-in. aperture to the temporal corner of *S*'s right eye from a distance of about $\frac{1}{2}$ in. The duration of the UCS was .10 sec., controlled by the action of a solenoid valve.

The CS was circular red light $1\frac{1}{4}$ in. in diameter produced by a Grason-Stadler Model E4580 multiple stimulus projector which was mounted equidistant from either side at a height of 70 in. in the cubicle wall facing *S*. The extra stimulus, or external inhibitor, was a 1,000-cps tone delivered at an intensity of approximately 70 db. (re .0002 dynes/cm^2) by an RCA beat frequency oscillator Model 154 through a $3\frac{1}{2}$-in. speaker mounted in the wall of the cubicle to *S*'s left.

The CS-UCS interval was 1 sec.; the offset of the CS coinciding with the onset of the UCS. On trials during which the tone was presented, its onset succeeded the onset of the CS by .50 sec., the two terminating together. All stimulus durations and interstimulus intervals were controlled by Grason-Stadler electronic interval timers. Intertrial intervals ranged from 12 to 28 sec. with a mean of 20 sec., and were controlled by a Gerbrands tape programmer whose action at the appropriate interval automatically initiated the sequence of events defining a trial.

Procedure.—After *S* was comfortably seated in the cubicle and the recording equipment attached, he was instructed in a manner designed to acquaint him with the general nature of the experimental procedure and to inhibit the production of voluntary responses by introducing a set to "respond naturally." All *S*s then received two nonreinforced CS presentations followed by three presentations of the UCS alone. Depending upon the experimental condition to which he was assigned, *S* then received 15, 30, or 60 paired presentations of the CS and UCS prior to receiving his first trial with the external inhibitor present. Group I received 60 CS-UCS trials, 1 trial with the external inhibitor and 5 more CS-UCS trials. Group II received 30 CS-UCS trials, an external inhibitor trial, 30 more CS-UCS trials, then a second external inhibitor trial followed by 5 CS-UCS trials. Group III *S*s received their first external inhibitor trial after 15 CS-UCS trials, their second after the thirtieth CS-UCS trial, and a third after the sixtieth CS-UCS trial, the latter again followed by 5 final CS-UCS trials. Thus provision was made to compare *S*s' performances following presentation of the external inhibitor after 15 and/or 30 reinforcements with the performance of *S*s who had not yet experienced the external inhibitor. All acquisition trials involving the extra stimulus were reinforced to prevent the development of discrimination and to permit observation of the effects, if any, of the external inhibitor upon the UCR.

The extinction procedure was identical for all *S*s. Three extinction trials involving only the CS were administered followed by 5 nonreinforced trials in which the tone was added to the CS as it had been in acquisition. This procedure was carried out with 57 *S*s, 19 from each group. The first *S* run in each condition was administered 10 straight extinction trials; subsequent addition of the tone produced no change in what was typical extinction behavior. This suggested that fewer nonreinforced presentations of the CS would be necessary to prevent the development of so much extinctive inhibition that no effects of the external inhibitor could be observed.

Subjects.—The *S*s were Duke University undergraduates who participated voluntarily as part of the requirements of the introductory psychology course. Usable data were collected from 60 of the 77 *S*s who participated; 9 of those remaining were discarded as voluntary responders, 6 failed to produce at least five conditioned responses during the acquisition period, and an exceedingly high resting blink rate made the records of the remaining 2 unreadable. The *S*s were assigned serially in order of their appearance at the laboratory to one of the three experimental groups. The assignment of *S*s to groups was partially restricted in order that each group would finally be composed of 14 males and 6 females, approximately the ratio of the sexes of the parent population.

RESULTS

Definition of the dependent variable.—A conditioned response was defined

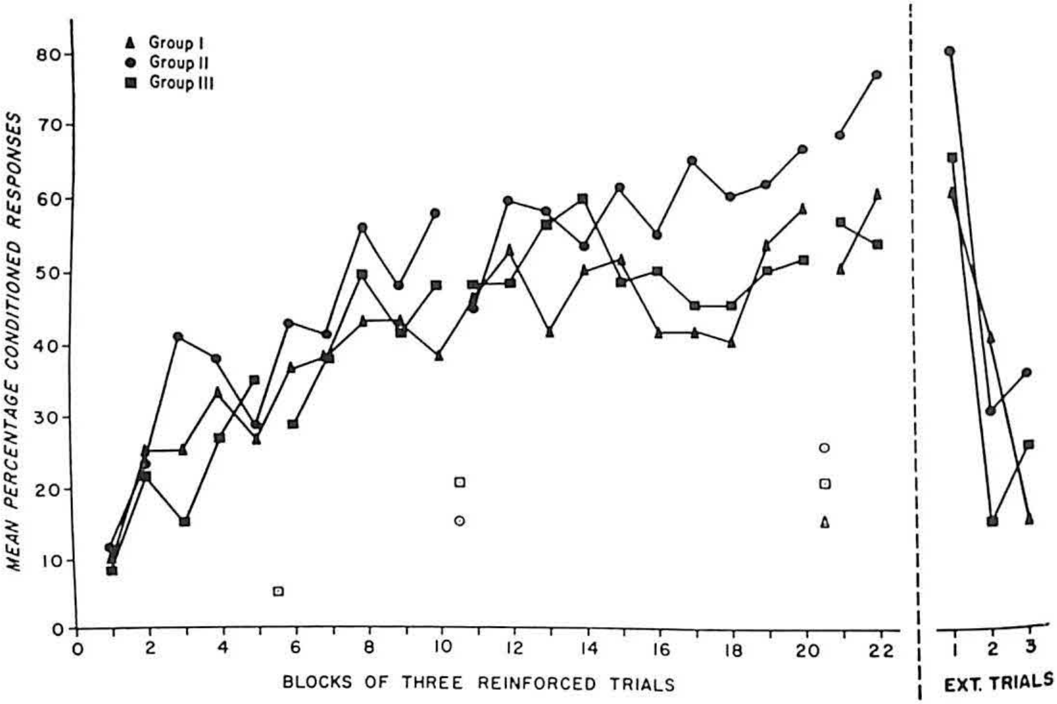


FIG. 1. Mean percentage of conditioned responses for the three groups presented by blocks of three trials in acquisition and trial by trial in extinction.

as any deflection from a stable baseline in the direction of closure that: (*a*) was greater than or equal to 1 mm.; (*b*) occurred at least 600 msec. after the onset of the CS; and (*c*) did not return to the baseline prior to the onset of the UCS. The rationale underlying the use of the latter criterion no doubt requires some elaboration.

It is well known that presentation of virtually any suprathreshold novel stimulus will evoke an orienting reflex on the part of a human *S* and that one component of this reflex is apt to be an eyeblink. Since, as Grant and Norris (1946) and others have shown, the latency of such a reflex blink may be as much as 110 msec., it was entirely possible in the present study for reflex (or alpha) blinks to the onset of the tone to be mistaken for conditioned responses on those trials in which the tone was intended to serve as an external inhibitor. By requiring that a conditioned response blend topographically with the unconditioned response, the likelihood of inadvertently including alpha responses on those trials was greatly reduced since the duration of an alpha reflex is not apt to be more than about 100 msec. As a check on the efficacy of this procedure, the data were tabulated with and without use of the blending criterion. On the reinforced trials that included the tone, the blending criterion eliminated 39.4% of the responses that would otherwise have been classified as CRs; it eliminated only 10.6% of such responses in the other acquisition trials. It is assumed that this difference reflects the presence of alpha reflexes to the onset of the tone which were appropriately omitted from the analysis. In order to assure statistical comparability, however, the blending criterion was applied to all reinforced responses thus making any statistical demonstration of external inhibition due to the presence of the tone somewhat more difficult than it would have been had the blending criterion only been applied to those responses made in the presence of the tone. The blending criterion was not used in the case of responses observed during extinction.

For each group, Fig. 1 shows the mean percentage of conditioned responses in acquisition (including the first extinction trial) plotted as a function of blocks of three trials. Performance in extinction is similarly plotted trial by trial for the three trials during which only the CS was presented. The open symbols in Fig. 1 represent the performance of the various groups on the trials which included the external inhibitor. Inspection of the figure clearly indicates that a large decrement in performance resulted from presentation of the external inhibitor but that the amount of decrement is not greatly affected either by the number of prior reinforcements or by the number of prior presentations of the external inhibitor itself. Statistical tests of all possible differences among these points confirm the latter impression.

The decrement which occurred in the presence of the external inhibitor was evaluated for each group separately by means of the Cochran *Q* test (Siegel, 1956); this technique simultaneously compared each *S*'s performance on the external inhibitor trial with his performance on each of the three immediately preceding acquisition trials. Similar analyses were performed on the pooled data of Groups II and III, following 30 CS-UCS trials and on the data of all three groups following 60 CS-UCS trials. In all cases, it was concluded that the decrement in performance due to the presentation of the external inhibitor was statistically reliable ($p < .05$).

Although analysis of the groups' performances on the external inhibitor trials reveals no differences which can be reliably attributed to differential habit strength as inferred from the number of prior reinforcements, examination of the data for individual *S*s provides evidence of a strong inverse relationship between the effectiveness of the tone as an inhibitor and level of conditioning.

The overall distribution of total number of conditioned responses produced in acquisition was dichotomized at the median; a 2×2 frequency table relating this index of response strength to whether or not a conditioned response was made on any of the external inhibitor trials exhibited a degree of association whose likelihood under the null hypothesis was found by Fisher's exact test to be .0358. Thus, support was provided for the Pavlovian contention that the effect of an extra stimulus presented during the interval of delay will be a function of the strength of the conditioned response being delayed; specifically, that a strong CR is harder to disrupt in this manner than a weak one.

Figure 2 is a series of tracings, showing the variety of effects produced by the introduction of the tone at the midpoint of the interval of delay. Although total inhibition of the observable CR occurred in 83.3% of the trials in which the tone was presented, it is of interest to examine some of the characteristics of the responses made on the remaining 16.7%. It might be, for example, that even if the tone were not completely effective as an inhibitor, partial effectiveness would serve to lengthen the latency of the

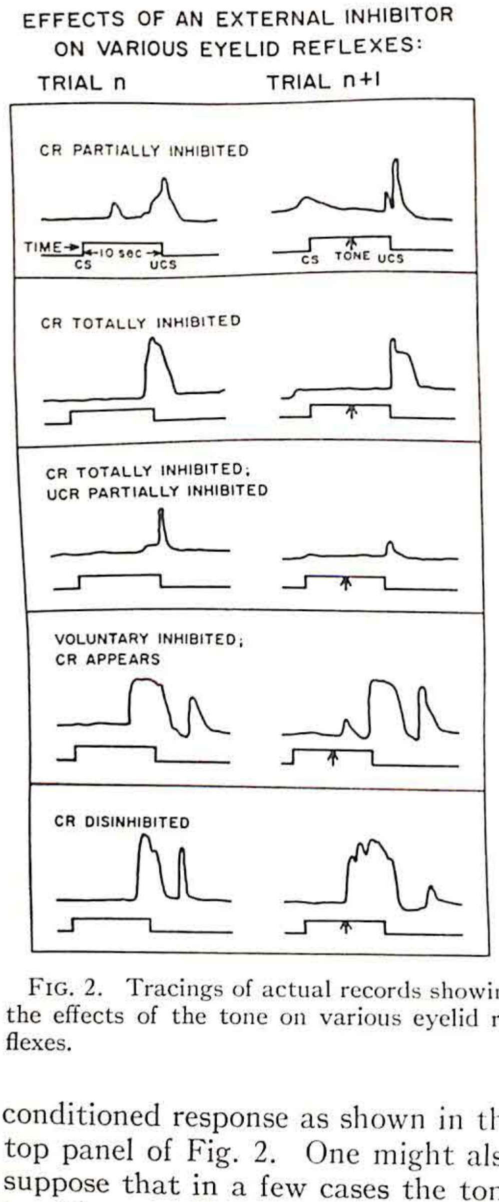


FIG. 2. Tracings of actual records showing the effects of the tone on various eyelid reflexes.

conditioned response as shown in the top panel of Fig. 2. One might also suppose that in a few cases the tone would operate as a disinhibitor, removing the inhibition of delay by decreasing the latency of the conditioned response. Such an effect is shown in the bottom panel of Fig. 2. In order to see if either or both of these effects occurred with any regularity in those cases where the tone failed to block completely the production of a conditioned response, response latencies were recorded for each *S* on the last external inhibitor trial on which a CR was produced. Also recorded were the latencies of the two immediately prior and two immediately subsequent conditioned responses. Analysis of these data revealed that introduction of the tone had no particular effect on the mean latency of either the responses made in the presence of the tone or those made immediately after its presentation. There is, however, a significant increase in the variability of these latencies. F ratios were formed between the variance of the CR latencies obtained on the trials with the external inhibitor present and the pooled variance of the latencies of the two previous conditioned responses, $F\ (29, 13) = 2.467, p < .025$; and between the pooled variance of the two subsequent conditioned responses and the pooled variance of the two previous conditioned responses, $F\ (29, 29) = 2.253,\ p < .05$. The comparison involving the latencies of the CRs in the presence of the external inhibitor and those of the two subsequent responses failed to reach significance ($F < 1.00$). Hence we have evidence that if the external inhibitor fails to produce total inhibition of the CR it may alter the latency not only of the conditioned response produced in its immediate presence but of the subsequent conditioned responses as well.

Extinction.—A Group × Trials analysis of variance was performed on the absolute numbers of CRs produced on the last five extinction trials by the 57 *S*s who received three nonreinforced presentations of the CS followed by five nonreinforced presentations of the CS plus the tone. The results of this analysis are summarized in Table 1.

One notes, by examining Table 1, that a nearly significant difference in performance on the last five extinction trials occurred as a function of the number of times the tone had been

TABLE 1

ANALYSIS OF VARIANCE OF CONDITIONED RESPONSES MADE WHEN THE TONE WAS ALSO PRESENT IN EXTINCTION

Source	*df*	*MS*	*F*
Groups (G)	2	5.067	2.924
Trials (T)	4	3.333	1.923
Linear	(1)	3.333	1.923
Quadratic	(1)	9.524	5.496*
G × T	8	1.733	
Total	14		

* $p < .05$.

presented and reinforced during acquisition; the mean numbers of CRs per *S* in Groups I, II, and III were .95, 1.26, and 1.47, respectively. This trend suggests that attempting to avoid the formation of discrimination during acquisition by reinforcing all presentations of the tone may have led to somewhat different levels of conditioning to the tone for the three groups; indirect evidence that a discrimination might indeed have been formed had the tone not been reinforced though it was presented at most only three times. Of greater interest, however, is evidence for the reliability of the quadratic nature of the function relating performance to trials. The CR totals for these five trials were, respectively, 15, 11, 12, 13, and 19. The increase of this function over the last three trials appears to confirm Wenger's finding that an external inhibitor in acquisition will, if presented during extinction, assume disinhibitory capacities. It must be added, however, that this interpretation of these results is somewhat tentative because of the absence of necessary controls against which to assess the effect of further ordinary extinction trials and the extent to which these responses may be UCRs to the tone.

DISCUSSION

The principal result of this study was its confirmation in human eyelid conditioning of the existence of a phenomenon of basic importance in classical conditioning: external inhibition. Specifically, it has been shown that the introduction of an extra stimulus during that portion of the interval of delay which is commonly thought to be excitatory will generally inhibit the production of a conditioned response.

The Pavlovian explanation for external inhibition is that a novel stimulus evokes an orienting reaction and that this reaction obliterates the conditioned response. The data support such an interpretation indirectly. It will be recalled that 39.4% of the responses made in the presence of the tone during acquisition was not considered in the analysis because they did not blend with the UCR whereas only 10.6% of the responses made in the absence of the tone was similarly eliminated. Inspection of the records indicates that the vast majority of the former was probably orienting reflexes; they occurred with appropriate latency and were usually of brief duration. In considering cases in which the tone failed to elicit an orienting blink but did inhibit the production of a conditioned response, it should be recognized that the blink reflex is probably only a part of the orienting reflex and perhaps not the essential part for the occurrence of external inhibition. Therefore, failure of the blink reflex to occur to all presentations of the tone cannot be taken as evidence that the orienting reaction also failed to occur.

The present demonstration of the susceptibility of the conditioned eyelid reflex to external inhibition and, tentatively, disinhibition, affords us the opportunity to examine at the human level the Pavlovian implications of these phenomena. An obvious next step, for example, would involve manipulating the intensity of the external inhibitor to see if the amount of external inhibition produced does vary directly with the intensity of the extra stimulus; the results of our preliminary study compared with those of the main study predict such an outcome since less inhibition was observed in the presence of a visual stimulus than in the presence of a more novel auditory one.

Manipulation of the temporal point of delivery of the extra stimulus may shed light on the role of the orienting reflex to the onset of the extra stimulus in the production of external inhibition; if external inhibition is nothing more than the result of the orienting reflex and its refractory period, extra stimuli presented early in the delay interval should have less of an inhibitory effect than those presented later. The possibility exists, however, that an extra stimulus presented at a point in the interval too late to effect the anticipatory CR would have an inhibitory effect on the UCR. The present study yielded no reliable evidence that a tone presented half way through the interval generally had any such effect on the UCR, but the effect was observed in isolated instances (see Fig. 2, third panel).

The results of the present study failed to confirm the expected presence of an induction effect in trials immediately following the presentation of the external inhibitor (see Fig. 1).[3] Since the preliminary study which suggested the likelihood of an inductive effect differed from the present one in that the external inhibitor was (*a*) visual and (*b*) presented *instead of* the CS rather than in conjunction with it, it seems likely that one or both of these differences are responsible for the disparity of the results. On the other hand, it may be that reinforcement delivered in the presence of a totally novel stimulus temporarily sensitizes *S* so that in the preliminary experiment, we observed a pseudo induction effect which was not, in fact, preceded by a great amount of internal inhibition. By contrast, in the present experiment, the external inhibitor apparently was not of sufficient intensity to produce an inductive effect but since it was presented in the presence of the CS subsequent reinforcement was probably not as sensitizing. Thus, it would appear that the demonstration of an induction effect following the presentation of an external inhibitor during the interval of delay must await the use of a more intense extraneous stimulus.

REFERENCES

Dufort, R. H., & Kimble, G. A. Ready signals and the effect of interpolated UCS presentations in eyelid conditioning. *J. exp. Psychol.*, 1958, **56**, 1–7.

Grant, D. A., & Norris, E. B. Eyelid conditioning as influenced by the presence of sensitized beta-responses. *J. exp. Psychol.*, 1946, **37**, 423–433.

Kimble, G. A., & Ost, J. W. P. A conditioned inhibitory process in eyelid conditioning. *J. exp. Psychol.*, 1961, **61**, 150–156.

Kimmel, H. D., & Pennypacker, H. S. Conditioned diminution of the UCR as a function of the number of reinforcements. *J. exp. Psychol.*, 1962, **64**, 20–24.

McAllister, W. R., & McAllister, D. E. The influence of the ready signal and unpaired UCS presentations on eyelid conditioning. *J. exp. Psychol.*, 1960, **60**, 30–35.

Pavlov, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.

Rodnick, E. H. Does the interval of delay of conditioned responses possess inhibitory properties? *J. exp. Psychol.*, 1937, **20**, 507–527.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Spence, K. W., & Runquist, W. N. Temporal effects of conditioned fear on the eyelid reflex. *J. exp. Psychol.*, 1958, **55**, 613–616.

Wenger, M. A. External inhibition and disinhibition produced by duplicate stimuli. *Amer. J. Psychol.*, 1936, **48**, 446–456.

(Received January 23, 1963)

---

[3] A *t* test comparing the performance of Group II on Trial Blocks 10 and 11 (separated by their first experience with the tone) yielded a ratio less than unity. Since this difference is one of the largest to be found in Fig. 1, either within or between groups on comparable or adjacent trial blocks, no effort was made to procure statistical support for either an induction effect or a locally perseverative inhibitory effect among the other points.

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 41–47

GENERALIZATION OF ACQUIRED FEAR AS A FUNCTION OF CS INTENSITY AND NUMBER OF ACQUISITION TRIALS[1]

OTELLO DESIDERATO

Connecticut College

28 rats received 3 or 25 trials of light CS and brief unavoidable shock. CS was bright for half the Ss and dim for the rest. A control group of 7 Ss received 25 backward conditioning trials with the bright CS. On subsequent generalization test trials, shock was omitted and latency of a hurdle jump response into an escape compartment was recorded, on presentation of test lights of varying intensities. Bright CS Ss were tested toward weaker, dim CS Ss toward stronger intensities. Obtained gradients of generalization of acquired fear, following conditioning with a bright CS, are approximately linear and tend to steepen with trials, on the 1st test day. Weak fear conditioning and, hence, no sloping gradients were obtained following backward conditioning or conditioning with a dim CS. Implications of these results for stimulus intensity dynamism and their relation to recent studies of acquired fear are discussed.

Following the initial study of Brown and Jacobs (1949), other investigators have used the method of prior fear conditioning to show that fear can be acquired and that its reduction will reinforce the learning of a new response (Goldstein, 1960; Gwinn, 1951; Kalish, 1954; McAllister & McAllister, 1962a). The procedure starts with a classical conditioning phase during which the animal receives a number of trials of CS paired with unavoidable shock of short, fixed duration. On subsequent trials, *S* learns to escape from the shock compartment by jumping a hurdle when the CS is presented alone; strength of fear acquired in the preceding phase is inferred from hurdle jump latencies across a series of test trials.

In the typical generalization of avoidance study, *S* must learn to make an avoidance response to the CS *prior* to generalization testing (Miller & Greene, 1954). In the present study, the aim was to obtain generalization gradients for a response never performed during the fear conditioning trials. Strength of acquired fear has been shown to vary with number of CS-UCS trials (Goldstein, 1960; Kalish, 1954), with intensity of UCS (Goldstein, 1960; Gwinn, 1951), and with delay between conditioning and test trials (McAllister & McAllister, 1962a). However, the generalization of fear acquired *prior* to the performance of the indicator response has not been reported. This study investigates the effect of CS intensity and number of CS-UCS trials, varied during fear conditioning, on the form of generalization gradients obtained on subsequent hurdle jumping trials.

---

[1] This research was supported by Grant MY-4705 from the National Institute of Mental Health, United States Public Health Service, and data were collected while the author was at Adelphi College. The author is grateful to Lois Rice, Mary Ann Bellia, and Frederick Arensberg for helping to collect the data and to Robert Kintz for his assistance in data analysis.

METHOD

Subjects

The *S*s were 35 experimentally naive male albino rats obtained from Blue Spruce Farms, Altamont, New York. The *S*s were 90–115 days old upon arrival and 130–155 days old at the start of the experiment. From the time of arrival and throughout the experiment all *S*s had Purina laboratory chow available as well as free access to water. Between time of arrival and the first experimental session, rats were handled for 5-min. periods every other day.

Apparatus

The apparatus consisted of two compartments separated by a guillotine-type door. The inside dimensions of the conditioning compartment were 13 in. long, $8\frac{1}{2}$ in. wide, and $7\frac{1}{2}$ in. deep. Three side panels were made of aluminum; the fourth side consisted of a wooden frame for the guillotine-type door. Panels, frame, and door were painted flat black. With the door open, the opening between compartments was $5\frac{1}{2}$ in. wide and $4\frac{3}{4}$ in. high. Escape from the conditioning compartment required climbing over a $1\frac{1}{2}$-in. hurdle consisting of the lower horizontal member of the wooden frame. One inch above the floor of the cage, the upper portion of the hurdle was beveled to a thin edge to discourage resting responses. When the door was lowered, its bottom edge rested on the beveled edge of the hurdle.

The compartment was covered with $\frac{1}{4}$-in. Plexiglas painted white except for an $8\frac{1}{2}$-in. circular area. The CS was light provided by a 60-w. frosted bulb in an $8\frac{1}{2}$-in. aluminum reflector which rested on top of this transparent area, facing down into the compartment. A variac was used to vary CS intensity and illumination was measured with a Photovolt meter, Model 200-A. All readings were taken with the photocell on the grid floor in the middle of the cage, facing upward toward the light source. On all acquisition trials, electronic timers timed CS and UCS presentations. The compartment floor consisted of $\frac{3}{16}$-in. steel rods, spaced $\frac{7}{16}$ in. apart, and wired to a shock generator and scrambler which supplied 1.2–1.8 ma. at 120 v. measured across the grids.

Inside dimensions of the escape compartment were 20 in. long, 7 in. wide, and 8 in. deep. This compartment was also painted flat black and had a hinged cover of clear Plexiglas. On generalization test trials, raising the door between the two compartments simultaneously started a Standard Electric timer and turned on the CS. Lowering the door after the animal had jumped the hurdle into the escape compartment automatically stopped the clock and turned off the CS.

Procedure

Design.—The general procedure was to expose the rat to a number of CS-UCS presentations in the conditioning compartment and then to record hurdle jump latencies in the presence of light stimuli of varying intensities. A mixed design was used with CS intensity and number of fear acquisition trials as the two independent treatments, and intensity of generalization test stimuli as the "within" effect. Twenty-eight *S*s were randomly assigned to a bright or dim CS intensity group. Each CS intensity group was further divided on the basis of number of fear acquisition trials: 3 or 25. The *S*s were run one at a time throughout the experiment.

Fear acquisition and generalization testing (Day 1).—On the first day, each *S* was permitted to explore the interior of the escape compartment for a 10-min. adaptation period. The *S*s were inserted directly into the compartment and the door to the conditioning compartment remained closed. The only illumination was a 10-w. ruby darkroom lamp in a ceiling fixture.

Immediately following adaptation, *S* was placed directly on the grid floor of the conditioning compartment. The door to the escape compartment remained closed. The *S*s in the 25 acquisition trial group immediately received their first fear acquisition trial consisting of 5 sec. of CS, with shock timed to begin 4 sec. after CS onset. The CS and UCS terminated simultaneously. Trials were repeated every 2 min., the twenty-fifth and last trial occurring 48 min. after the introduction of *S* into the conditioning compartment. In order to equate the total amount of time all *S*s spent in the presence of intracompartment cues other than the CS, *S*s in the 3 acquisition trial group did not receive their first fear acquisition trial until 44 min. had been spent in the box. Thus, their third and last trial also arrived after 48 min. in the compartment. Bright CS intensity was 125 ft-c; the intensity of the dim CS was .5 ft-c.

Two minutes following the termination of the last fear acquisition trial, every *S* was returned to his home cage in a remote corner of the room for a 90-min. period during which another animal was conditioned. During this rest period, *S*'s cage was completely covered with a heavy black cloth to maintain dark adaptation. Food and water were available.

Generalization testing began 90 min. after the last fear acquisition trial. On all acquisition trials, the four inside panels of the conditioning box had been covered with a layer of white cardboard. These were now removed to expose the black frame and door, and the black surfaces of the three remaining compartment walls. To further attenuate the fear inducing effect of compartment cues, the grid floor was completely covered with a rectangular section of $\frac{1}{8}$-in. plywood, painted flat black.

On each test trial, *S* was placed into the conditioning compartment and, after 15–30 sec., the door to the escape compartment was raised. When the door had traveled upward 2 in., a microswitch simultaneously activated CS and timer. When the rat cleared the hurdle, the door was immediately lowered, turning off CS and stopping the timer. A rubber stop prevented lowering the door all the way, leaving a $\frac{3}{8}$-in. gap at the bottom for the rat's tail (thereby encouraging *E* to shorter and less variable personal latencies). The rat remained 15–25 sec. in the escape compartment and was then returned to his home cage for 2 min. before the start of the next trial. If the hurdle was not crossed within 60 sec. of the onset of the CS, the door was lowered and, after 15 sec., the rat was returned to his home cage for the usual intertrial period. On such failures to respond, latencies of 60 sec. were recorded. Shock was never administered on test trials.

On Day 1, each *S* received 18 test trials, 6 trials to each of three light intensities: 125, 7.9, and .5 ft-c. These values are equal log units apart. Order of test stimuli was random within blocks of three.

Second generalization testing (*Day 2*).—Eighteen test trials were again administered on the next day, following the same procedure used on Day 1 generalization test trials.

Control group.—In order to determine whether latency differences reflected differences in the noxious properties of varying light intensities, rather than amount of generalization, a control group consisting of seven *Ss* was given 25 backward conditioning trials with the bright CS. All conditions were the same as for the corresponding experimental group, except that CS onset occurred 20 sec. *after* UCS onset. This group was also intended to show whether fear conditioning had occurred to compartment cues.

TABLE 1

ANALYSIS OF VARIANCE OF RECIPROCAL LATENCIES

Source	*df*	*F*
Between *Ss*	27	
Acquisition trials (T)	1	1.51
CS intensity (CS)	1	1.63
T × CS	1	
Error (b)	24	(.0888)[a]
Within *Ss*	120	
Test stimulus intensity (S)	2	7.59**
S × T	2	4.59*
S × CS	2	5.48**
S × T × CS	2	1.44
Error (S)	48	(.0027)[a]
Days (D)	1	3.54
D × T	1	
D × CS	1	
D × T × CS	1	
Error (D)	24	(.0189)[a]
S × D	2	2.00
S × D × T	2	
S × D × CS	2	1.00
S × D × T × CS	2	
Error (S × D)	48	(.0010)[a]
Total	167	

[a] Error estimate.
* $p < .05$.
** $p < .01$.

RESULTS

On each test day, every *S* received six trials to each of three light intensities. Each of the six latencies was converted to its reciprocal, and a mean reciprocal latency for each light intensity was then computed for every *S*. This score became the unit employed in the analysis of variance.

Table 1 summarizes the analysis of variance of reciprocal latencies. The significant S × T interaction indicates that the gradients for the 3 and 25 acquisition trial groups diverged. However, Fig. 1 shows that a falling gradient is limited to the group receiving 25 acquisition trials. Means for test stimuli within this group differed significantly, $F(2,48) = 11.58$, $p < .01$. Mean differences for the 3 acquisition trial group were not significant. In both cases, the error term used was the within-stimuli residual (Table 1).

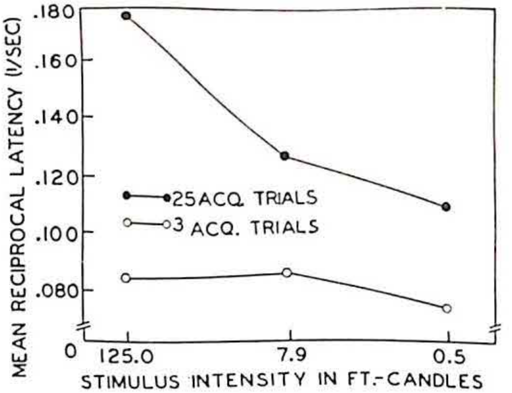


FIG. 1. Mean reciprocal latencies of the hurdle jump response to three test stimulus intensities as a function of number of fear acquisition trials.

Comparison of means at each separate test stimulus revealed that latencies for the 25 acquisition trial group were significantly shorter than for the 3 acquisition trial group only at the level of the brightest generalization test stimulus, F (1, 54) = 5.09, $p < .05$. The hypothesis of no difference between means at the intermediate and dim test stimulus intensities could not be rejected, F (1, 54) = 1.30, $p > .25$; F (1, 54) = 1.23, $p > .25$. The error term was the within-*S*s residual, computed separately for each test stimulus intensity.

Table 1 also shows a significant S × CS interaction, indicating a difference in the form of the generalization curve for *S*s conditioned to the bright as opposed to the dim CS. Figure 2 shows a steep gradient for the bright CS group, F (2, 48) = 12.33, $p < .01$; for the dim CS group, the gradient was relatively flat. In the latter group, the mean square for test stimuli was smaller than the error mean square. In both cases, the within-stimuli residual (Table 1) was the error term employed. The bright CS group gave significantly shorter latencies than the dim CS group when tested at the brightest stimulus intensity, F (1, 54) = 8.56, $p < .01$, as well as at the intermediate intensity, F (1, 54) = 6.93, $p < .05$. The difference at the dimmest test intensity was not significant. For all comparisons the within-*S*s residual, computed separately for each test stimulus level, was the error term.

To examine the possibility that aversive effects of strong illumination may have accounted for the steep gradient shown by the bright CS group, gradients for the control group and the 25 trial, bright CS group were compared. Figure 3 shows that the control group's gradient was essentially flat. A separate analysis of variance applied to these two groups is summarized in Table 2. Although the main effect for vertical displacement was not significant, the significant S × E/C interaction indicates that light intensity, per se, was not responsible for producing the sloping curves shown in Fig. 1 and 2. The long latencies obtained under backward conditioning procedures also demonstrate the failure of intracompartment cues *other than the CS* to evoke hurdle jumping to any appreciable degree.

Having established the occurrence of stimulus generalization for *S*s con-

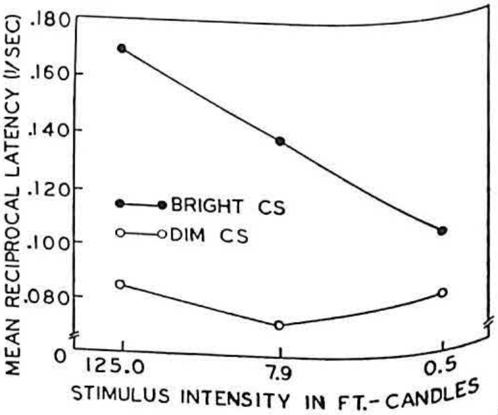


FIG. 2. Generalization gradients obtained after conditioning the fear reaction to a bright or dim light CS.

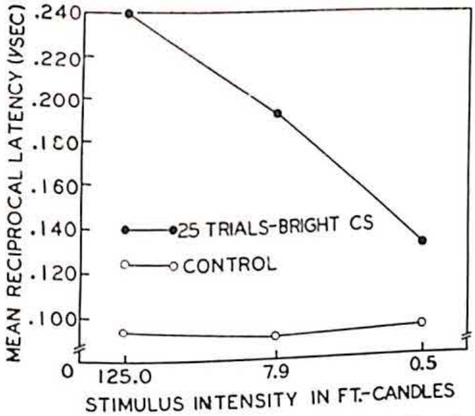


FIG. 3. Comparison of generalization gradients for experimental and control groups. (Both groups received 25 trials of bright CS and shock, the only difference being the backward conditioning procedure used for the control group.)

ditioned to a bright CS, data for these *S*s were then examined for evidence of systematic changes in the form of the gradients within a daily session. Accordingly, the six trials to each test stimulus given each *S* on Day 1 were subdivided into three blocks of two trials each. The mean of each two-trial block became *S*'s raw score for that block. Data for Day 1 were arranged in a 2 × 3 × 3 mixed design and an analysis of variance was performed; number of acquisition trials (3 vs. 25) constituted the between variable, while test stimulus intensity and test trials comprised the within variables. The analysis was repeated for data collected on Day 2. Table 3 shows the S × B interaction to be significant for Day 1 but not for Day 2. It is evident in Fig. 4 that, on Day 1, the gradient began fairly flat on the first trials but steepened sharply by the last trial block. Separate analysis of each gradient revealed that mean latencies to test stimuli differed significantly only on Trials 5–6, F (2, 24) = 3.94, $p < .05$.

On Day 2, the F ratio for test stimuli indicates a significant generalization effect ($p < .001$), but the form of the gradient no longer varied with trial blocks. Although Fig. 4 shows a flatter gradient for the first than for the later trial blocks, the

TABLE 2

ANALYSIS OF VARIANCE COMPARING 25 TRIAL, BRIGHT CS GROUP WITH CONTROL GROUP

Source	*df*	*F*
Between *S*s	13	
Experimental vs. control (E/C)	1	2.27
Error (b)	12	(.0789)[a]
Within *S*s	80	
Test stimulus intensity (S)	2	6.39**
S × E/C	2	5.94**
Error (S)	24	(.0033)[a]
Days (D)	1	1.72
D × E/C	1	
Error (D)	12	(.0195)[a]
S × D	2	1.67
S × D × E/C	2	
Error (S × D)	24	(.0009)[a]
Total	83	

[a] Error estimate.
** $p < .01$.

TABLE 3

SEPARATE ANALYSES OF VARIANCE FOR THE BRIGHT CS GROUP ON DAYS 1 AND 2

Source	*df*	Day 1	Day 2
		F	*F*
Between *S*s	13		
Acquisition trials (T)	1	2.00	1.22
Error (b)	12	(.1905)[a]	(.1976)[a]
Within *S*s	112		
Test stimulus intensity (S)	2	2.76	14.76**
S × T	2	1.74	2.58
Error (S)	24	(.0089)[a]	(.0045)[a]
Trial blocks (B)	2	1.15	1.90
B × T	2		
Error (B)	24	(.0168)[a]	(.0115)[a]
S × B	4	2.70*	1.72
S × B × T	4	2.00	
Error (S × B)	48	(.0030)[a]	(.0079)[a]
Total	125		

[a] Error estimate.
* $p < .05$.
** $p < .01$.

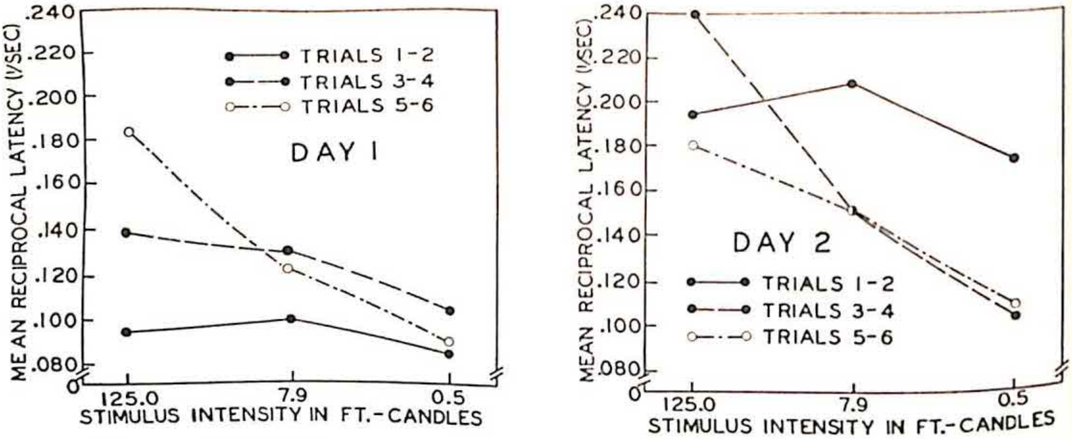


FIG. 4. Generalization gradients as a function of test trials, following 25 fear acquisition trials on which a bright CS was used.

interaction mean square was not significant.

DISCUSSION

No evidence of generalization was obtained when the number of fear conditioning trials was small or when the CS intensity was dim. The most obvious interpretation of this result is that, under these conditions, no fear conditioning to the CS takes place. The long latencies shown in Fig. 1 and 2 are comparable to the latency scores of the control group in Fig. 3.

Conditioning to the CS clearly occurs when both a bright CS and many acquisition trials are employed, as shown by the mean latency to the CS in Fig. 3. These conditions produced a generalization gradient which is essentially linear when plotted against stimulus intensity measured in equal log units. The steep slope is in agreement with the findings of Brown (1942) and Miller and Greene (1954) for *S*s trained at high and tested at lower intensities. The failure of the control group to show faster responding to the brighter stimuli indicates that the gradient shown by the experimental group cannot be explained on the basis of aversive properties of bright stimuli.

The extent to which stimulus intensity dynamism affected the gradients shown in Fig. 2 is an interesting question. The low, flat gradients shown in Fig. 1, 2, and 3 do not suggest the operation of stimulus intensity dynamism. These gradients do suggest, however, that the CS does not become an effective cue for hurdle jumping under conditions of (*a*) few acquisition trials, (*b*) low intensity, or (*c*) backward conditioning (control group). Hull's (1952) stimulus dynamism postulate may thus not apply to cues which have failed to become conditioned stimuli. If increasing stimulus intensity produces dynamogenic effects only after a cue has been conditioned, then the steep gradients shown in Fig. 2 and 3 may reflect both generalization and dynamism effects.

The frequently reported steepening of the gradient with increasing test trials (Brown, 1942; Wickens, Schroder, & Snide, 1954) occurred on Day 1. However, comparison with the usual extinction-type procedure for generalization testing is complicated by the fact that, in the present study, secondary reinforcement in the form of fear reduction operated throughout the test trials, presumably extinguishing slowly as the trials progressed. Moreover, whereas in the typical situation the CR is at its strongest at the beginning of testing and gradually weakens, the hurdle jump response is weak at first, rapidly rises to a maximum, and levels off or begins to extinguish on the second day of testing (Goldstein, 1960; Kalish, 1954; McAllister & McAllister, 1962a, 1962b). Figure 4 suggests that, on

Day 1, the effectiveness of the original CS in evoking hurdle jumping increases with the number of reinforced responses; as the effectiveness of the CS increases, the generalization gradient has an opportunity to emerge. By the second day, the response to the CS remains at approximately maximum strength and, therefore, no within-session gradient changes are evident.

Recent studies by McAllister and McAllister (1962b) have shown that the hurdle jump response may become conditioned to apparatus cues as well as to the CS. Such conditioning would account for their finding and Goldstein's (1960) that the hurdle jump response can be learned even if the backward conditioning paradigm is used in the fear conditioning trials. The lack of evidence for backward conditioning in the present study may be due to our deliberate effort to change apparatus cues present during the fear conditioning and the hurdle jumping test phases. In the McAllister and Goldstein studies, although different compartments were used for the conditioning and testing periods, they were constructed to appear identical. In the present study, the grid floor was covered with a wooden panel and the compartment walls were changed from white to black. These changes in visual and tactual apparatus cues were apparently sufficient to preclude conditioning of the fear reaction to cues present during the test trials.

Of some interest is the generally weak conditioning obtained in the present study, even under optimal conditions (bright CS, many training trials). Recent evidence reported by McAllister and McAllister (1962a) indicates that learning of the indicator response requires some delay between conditioning and hurdle jump tests. In their study, learning occurred after a 1-day delay but not after a 2-min. delay. Studies reporting better conditioning than obtained in the present study usually involve some delay between conditioning and hurdle jumping (Goldstein, 1960; Kalish, 1954; Kent, Wagner, & Gannon, 1960). The 90-min. delay employed in the present study may have been only partially sufficient (*a*) to permit dissipation of a "general emotional state" which might interfere with hurdle jumping (McAllister & McAllister, 1962a) or (*b*) for fear to "incubate" to an appreciable level of strength (Bindra & Cameron, 1953).

REFERENCES

Bindra, D., & Cameron, L. Changes in experimentally produced anxiety with the passage of time: Incubation effect. *J. exp. Psychol.*, 1953, **45**, 197–203.

Brown, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.

Brown, J. S., & Jacobs, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.

Goldstein, M. L. Acquired drive strength as a joint function of shock intensity and number of acquisition trials. *J. exp. Psychol.*, 1960, **60**, 349–358.

Gwinn, G. T. Resistance to extinction of learned fear-drives. *J. exp. Psychol.*, 1951, **42**, 6–12.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Kalish, H. I. Strength of fear as a function of the number of acquisition and extinction trials. *J. exp. Psychol.*, 1954, **47**, 1–9.

Kent, N. D., Wagner, M. K., & Gannon, D. R. Effect of unconditioned response restriction on subsequent acquisition of a habit motivated by "fear." *Psychol. Rep.*, 1960, **6**, 335–338.

McAllister, W. R., & McAllister, D. E. Postconditioning delay and intensity of shock as factors in the measurement of acquired fear. *J. exp. Psychol.*, 1962, **64**, 110–116. (a)

McAllister, W. R., & McAllister, D. E. Role of the CS and of apparatus cues in the measurement of acquired fear. *Psychol. Rep.*, 1962, **11**, 749–756. (b)

Miller, W. C., & Greene, J. E. Generalization of an avoidance response to varying intensities of sound. *J. comp. physiol. Psychol.*, 1954, **47**, 136–139.

Wickens, D. D., Schroder, H. M., & Snide, J. D. Primary stimulus generalization of the GSR under two conditions. *J. exp. Psychol.*, 1954, **47**, 52–56.

(Received January 28, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 48–51

DISCRIMINATION LEARNING UNDER VARIOUS COMBINATIONS OF FOOD AND SHOCK FOR "CORRECT" AND "INCORRECT" RESPONSES[1]

GEORGE J. WISCHNER, RICHARD C. HALL, AND HARRY FOWLER

University of Pittsburgh

To assess various interpretations of the paradoxical facilitating effect of shock for the correct response on visual discrimination learning, the performances of shock-right, shock-both, and no-shock groups were investigated employing a noncorrection procedure; a novel food-both group received both shock and food for "correct" responses and only food for "incorrect" responses. Efficiency of performance, beginning with the poorest group, was food-both, shock-right, shock-both, and no-shock, the difference between any 2 groups being statistically significant except for comparison of the shock-both and no-shock groups. The data support an avoidance rather than a cognitive-perceptual interpretation of the effect of shock and delimit any generalization of a facilitating effect of shock for the correct response.

The present study was designed to assess alternative interpretations which have been offered to account for the paradoxical facilitating effect on discrimination performance of shock for the *correct* response. These interpretations have included the following: (*a*) punishment serves to slow down the animal at a choice point, making it more sensitive to the discriminanda (Muenzinger, 1934; Muenzinger, Bernstone, & Richards, 1938); (*b*) punishment may be helpful as an "emphasizer" or detrimental as a "distractor" depending on the way it influences the learner's perceptions and cognitions (Hilgard, 1956; Tolman, Hall, & Bretnall, 1932); and (*c*) punishment, serving as a differential cue, may act as a secondary reinforcer (Freeburne & Taylor, 1952).

In an attempt to assess the last interpretation, Freeburne and Taylor (1952) trained rats on a discrimination task which involved shock for both correct and incorrect responses. The superior performance of this shock-both group, relative to a no-shock group, was interpreted as supporting a "sensitizing" function of shock. However, Prince (1956) questioned these findings on certain methodological grounds, and, in his study, obtained no difference between a shock-both and no-shock group.

Because of the equivocal findings relating to the shock-both condition, and other data which indicate that shock for the correct response retards performance (Wischner, 1947), performance retardation being greater with stronger shock intensities (Wischner, Fowler, & Kushnick, 1963), the present study attempted to reassess the performances of shock-right and shock-both groups. Also considered was a new, food-both group which received shock for the "correct" response but food for *both* "correct" and "incorrect" responses. According to the cognitive-perceptual and sensitizing interpretations, all three shock groups should show equally superior performance relative to no-shock controls.

---

[1] This study was initiated by Richard C. Hall during his tenure as a National Science Foundation Predoctoral Fellow and was completed with partial support from a National Science Foundation grant, No. G-14312, to G. J. Wischner and H. Fowler.

METHOD

Subjects

The *S*s were 28 naive, male albino rats of the Sprague-Dawley strain, about 80–100 days old at the start of the experiment. The *S*s were maintained in individual cages in the experimental room under controlled temperature and an artificially illuminated day-night cycle.

Apparatus

The apparatus was an enclosed T maze, with a stem 20 in. long, each arm 22 in. long, and each goal compartment 11 in. square. The height of the maze was uniformly 4 in. and the width of the stem and arms 2½ in. The initial 8-in. portion of the stem was separated from the remaining portion by a guillotine door and served as the start compartment. Guillotine doors were also positioned 3 in. from the center of the choice point in each arm and at the goal end of each arm.

The stem and arms were constructed of two L shaped strips of galvanized sheet metal. One L shaped strip served as one wall and half of the floor; the other L served as the other half of the floor and the other wall, with the two strips being separated at the floor by a ½-in. gap. Each goal compartment had a galvanized sheet metal floor and Plexiglas walls painted white except for the far wall. A frosted bulb (7 w., 120 v.), enclosed in a metal reflector, was located outside and directly behind the far wall of each goal; in front of the wall, there was a clear Plexiglas food well, 1 in. in diameter. Clear Plexiglas also served as the top of each maze section.

A matched-impedance shocking system delivered a single 60-cps ac shock for .2 sec. through a .25-megohm resistor to the two metal strips comprising the maze arms. Thus, *S* received shock as long as its feet were touching both halves of the floor. The narrowness of the arms prevented *S* from running along only one side of the floor. The shock could be administered to *S* in either maze arm by *E*'s operation of a microswitch.

Procedure

The procedure included both pretraining and training phases. One week prior to pretraining, *S*s were started and maintained for the duration of the experiment on a daily diet of 12 gm. Purina lab checkers, with water available ad libitum. During this week, each *S* was handled daily for about 5 min.

During the pretraining and training trials, *S*s were kept in individual compartments of a hardware cloth detention cage. Prior to *S*'s daily run in the maze, which took place during the dark phase of its day-night cycle, *S* received about 5 min. adaptation to the illumination conditions of the experimental room. During the pretraining trials, *S* was run only under general, overhead room illumination; during the training trials, when one or the other of the goal lights was on, the only other source of illumination was a 15-w. flourescent lamp positioned 2 ft. above and 3 ft. behind the choice point of the maze.

Pretraining was administered to reduce possible position preferences and to habituate *S* to the apparatus. No shock was administered during this phase. For pretraining, each *S* received a total of 10, forced, reinforced trials, 5 per day, randomly distributed with the restriction that half of the trials were to the left goal and half to the right. Forcing to one or the other goal was accomplished by lowering one of the guillotine doors at the choice point. The reinforcement provided per trial was a single P. J. Noyes sucrose pellet (4 mm., 45 mg.). On each trial *S* was detained in a goal compartment for 1 min. and then returned to the start compartment for the next trial, the interval between successive trials within a day being about 30 sec.

Training began on the day following termination of pretraining. For training, the 28 *S*s were randomly assigned to four groups of 7 *S*s each. During training, all *S*s received free-choice trials with one or the other of the goal compartment bulbs being illuminated on each trial. On each trial, *S*s of three groups received both food and shock for selection of the lighted "correct" goal; however, for selection of the unlighted, "incorrect" goal, one of these groups received only food, another only shock, and the third neither food nor shock. These groups will be referred to as the food-both group (SF-F), the shock-both group (SF-S), and the shock-right group (SF-O), respectively. The fourth group (F-O), a no-shock control, received only food for selection of the lighted alley and neither food nor shock for selection of the unlighted alley.

All groups received 5 trials per day for the first 20 training trials and then 10 trials per day thereafter. Throughout training

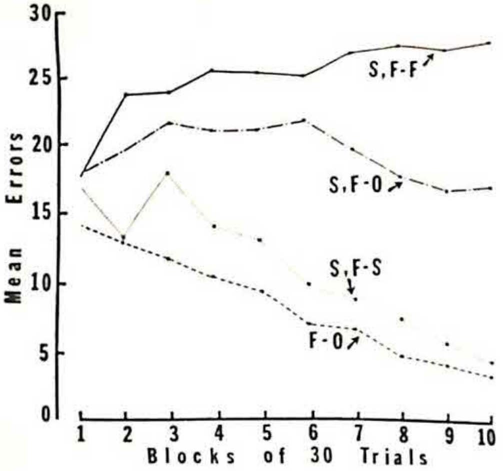


FIG. 1. Group mean errors over blocks of 30 trials.

and for all *Ss*, the right-left position of the lighted goal was LRRLL RLRRL for even 10s of trials, and LRRLR LLRRL for odd 10s of trials. On any training trial, food reinforcement was a single P. J. Noyes sucrose pellet (4 mm., 45 mg.), detention time in a goal 1 min., and the interval between successive trials within a day, about 30 sec. During Trials 1–20, the intensity of shock was 58 v. as measured across the output of the transformer; thereafter, the intensity was increased to and maintained at 65 v. for the remaining trials. Shock was administered when *S* reached a point midway down a maze arm.

Each *S* was run until it met a criterion of 19 "correct" choices on 20 consecutive trials or until a total of 300 trials had been completed. Since food was available for the SF-F group in both goals, the criterion of learning for these *Ss* was 19 choices on 20 consecutive trials of either the lighted or the unlighted alley, but not both.

RESULTS AND DISCUSSION

Group mean errors, i.e., responses to the unlighted goal, are presented in Fig. 1 in blocks of 30 trials over the 300 trials of the experiment. Early in training, mean errors for the several groups were at a chance level of performance. With training, however, mean errors increased for the SF-F group, remained relatively stable for the SF-O group, and decreased for both the SF-S and F-O groups. As indicated by an analysis of variance of group mean errors over the entire 300 training trials, these performance differences were highly reliable, $F(3, 24) = 24.9$, $p < .001$. Further, Duncan's multiple-range test shows that the difference between any two groups was statistically significant ($p < .05$ in every case) except in the comparison of the F-O and SF-S groups.

These findings seem to be more in accord with an avoidance rather than a perceptual-sensitization interpretation of the effect of shock. According to the avoidance conception, *S* learns to avoid those cues associated with shock (e.g., Wischner et al., 1963). Thus, for the shock-both (SF-S) group, the experience of shock in both the lighted and unlighted alleys led to a comparable avoidance of both alleys, with the food reinforcement in the lighted alley being the primary determinant of choice. For both the shock-right (SF-O) and food-both (SF-F) groups, on the other hand, shock administered only for response to the lighted goal led to an avoidance of that goal, with selection of the unlighted alley by the SF-F group being augmented presumably by the food reinforcement also obtained in that alley. Thus, by the end of training, the SF-F *Ss* were responding solely to the nonshock, unlighted alley.

One might argue that a cognitive-perceptual or sensitizing interpretation of the present data is also tenable, were it assumed that the shock intensity of 58–65 v. employed in the present study is so intense as to produce "perceptual distraction" rather than "emphasis." Thus, the inferior performances of the shock-right (SF-O) and food-both (SF-F) groups would be expected. It seems, however, that this interpretation also requires a marked disruption in the performance of the shock-both (SF-S) group. The lack of a significant difference in performance between the SF-S and F-O groups, together with the comparable data reported by Prince (1956) for these groups, suggests that the cognitive-perceptual interpretation is not altogether tenable. Moreover, it would appear that the shock experience provided *S* in the present study was not intense, but relatively mild, considering the fact that the shock stimulus was always a brief .2-sec. duration, as contrasted with the uncontrolled and varying shock durations typically employed in other studies. Also, with the matched-impedence shock system used, *S*'s threshold of reaction to shock is obtained at about 30–40 v. (cf. Campbell & Teghtsoonian, 1958).

It is concluded that the present findings, the comparable data reported by Prince (1956), by Wischner (1947), and more recently by Wischner, Fowler, and Kushnick (1963) support strongly the avoidance interpretation of the effect of shock in discrimination learning, thereby delimiting quite clearly any generalization of a facilitating effect of shock for the correct response.

REFERENCES

Campbell, B. A., & Teghtsoonian, R. Electrical and behavioral effects of different types of shock stimuli on the rat. *J. comp. physiol. Psychol.*, 1958, **51**, 185–192.

Freeburne, C. M., & Taylor, J. E. Discrimination learning with shock for right and wrong responses in the same subject. *J. comp. physiol. Psychol.*, 1952, **45**, 264–268.

Hilgard, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1956.

Muenzinger, K. F. Motivation in learning: I. Electric shock for correct responses in the visual discrimination habit. *J. comp. Psychol.*, 1934, **17**, 439–448.

Muenzinger, K. F., Bernstone, A. H., & Richards, L. Motivation in learning: VIII. Equal amounts of electric shock for right and wrong responses in a visual discrimination habit. *J. comp. Psychol.*, 1938, **26**, 177–186.

Prince, A. I., Jr. Effect of punishment on visual discrimination learning. *J. exp. Psychol.*, 1956, **52**, 381–385.

Tolman, E. C., Hall, C. S., & Bretnall, E. P. A disproof of the law of effect and a substitution of the laws of emphasis, motivation, and disruption. *J. exp. Psychol.*, 1932, **15**, 601–614.

Wischner, G. J. The effect of punishment on discrimination learning in a non-correction situation. *J. exp. Psychol.*, 1947, **37**, 271–284.

Wischner, G. J., Fowler, H., & Kushnick, S. A. Effect of strength of punishment for "correct" or "incorrect" responses on visual discrimination performance. *J. exp. Psychol.*, 1963, **65**, 131–138.

(Received January 29, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 52–58

A TEST OF THREE MODELS FOR STIMULUS COMPOUNDING WITH CHILDREN[1]

R. C. ATKINSON, R. C. CALFEE
Stanford University

G. R. SOMMER, W. E. JEFFREY, AND R. SHOEMAKER
University of California, Los Angeles

An experiment on stimulus compounding was conducted using 5th- and 6th-grade students as Ss. On each of a series of 960 discrete trials, S was presented with 1 of 3 stimuli, s_1, s_2, or s_3, or with a compound of these stimuli, $(s_1 + s_2)$, $(s_1 + s_3)$, $(s_2 + s_3)$, or $(s_1 + s_2 + s_3)$. The S responded to each stimulus presentation by selecting 1 of 2 response keys, A_1 or A_2. If π_i is the probability that A_1 is reinforced when s_i is presented then $\pi_i = 1.0$ and $\pi_2 = .33$ for all Ss. 3 experimental groups were differentiated by π_3 which took the values .25, .50, and .75. No reinforcement event followed the presentation of a compound. 3 models are described that generate predictions for response probabilities to a compound stimulus in terms of response probabilities to the component stimuli. Model I postulates a simple averaging rule. Model II assumes that the influence of each component stimulus is proportional to the likelihood that it elicits a correct response. Model III assumes that the response to the compound is determined solely by the one component that is most frequently reinforced. It was found that Model II gave the best account of these data.

A central issue in many current theories of learning deals with the problem of predicting behavior in the presence of a new stimulus compound that is constructed by combining component stimuli on which *S* has had previous discrimination training. As an example of the type of problem we have in mind, consider a situation where on each trial *S* is required to make either an A_1 or an A_2 response. If, after training, he tends to make A_1 with probability p_t on trials when a tone is presented, and with probability p_l when a light is presented, then what will be the probability of an A_1 response when light and tone are presented simultaneously. Obviously the probability p_{tl} of making A_1 to the compound stimulus (tone + light) will be at least partially dependent on the values of p_t and p_l. In terms of general psychological considerations there are several plausible functions relating p_{tl} to p_l and p_t; the purpose of this paper is to determine which of these functions provides the best account of data collected in a prediction experiment using young children.

The experimental situation involves a series of 960 discrete trials for each *S*. The trials are of two types: learning trials and test trials. Learning trials are initiated with the presentation of one of three comparable but clearly distinct stimuli; these stimuli are denoted s_1, s_2, and s_3. The *S* is required to make one of two responses (A_1 or A_2) and the trial is then terminated with either an E_1 or E_2 reinforcing event. The occurrence of E_i means that A_i was the correct response for that trial. The schedule for presenting reinforcing events on learning trials is specified by the parameter π_i; when s_i

---

[1] This research was supported by the National Science Foundation (Grants 10041 and 24264) and by the National Institute of Mental Health (Grant MH-05184).

is presented an E_1 occurs with probability π_i and an E_2 with probability $1 - \pi_i$. Interspersed among the learning trials are test trials. On test trials one of the following four stimulus compounds is presented: $(s_1 + s_2)$, $(s_1 + s_3)$, $(s_2 + s_3)$, or $(s_1 + s_2 + s_3)$. The S is required to make an A_1 or A_2 response to the presentation of the compound, but no reinforcing event is given.

We shall consider three models that yield predictions for responses in the presence of a stimulus compound. Let p_i denote the probability of an A_1 response to the presentation of s_i; p_{ij} the probability of A_1 to the presentation of the compound $(s_i + s_j)$; and p_{123} the probability of A_1 to the compound $(s_1+s_2+s_3)$. The three models to be considered are as follows:

Model I

$$p_{ij} = \omega[\tfrac{1}{2}(p_i + p_j)] + (1-\omega)\tfrac{1}{2}, \qquad [1]$$

$$p_{123} = \omega[\tfrac{1}{3}(p_1 + p_2 + p_3)] + (1-\omega)\tfrac{1}{2}.$$

This hypothesis is derived from models of discrimination learning proposed by Schoeffler (1954), Burke and Estes (1957), Atkinson and Estes (1963), and others. The first part of the right hand side of the equation states that each component stimulus contributes *equally* in determining the response to the compound; i.e., the probability of an A_1 response to the compound $(s_1 + s_2)$ is simply the average of p_1 and p_2. The second part of the equation allows for the possibility that there may be some regression toward random responding in the presence of a new stimulus compound. The parameter ω weights these two effects and we assume that the value of ω is the same for all stimulus compounds.

Model II

$$p_{ij} = \omega\left[\frac{c_i}{c_i + c_j}\,p_i + \frac{c_j}{c_i + c_j}\,p_j\right] + (1-\omega)\tfrac{1}{2}, \qquad [2]$$

$$p_{123} = \omega\left[\frac{c_1p_1 + c_2p_2 + c_3p_3}{c_1 + c_2 + c_3}\right] + (1-\omega)\tfrac{1}{2},$$

where $c_i = p_i\pi_i + (1 - p_i)(1 - \pi_i)$ and is the probability of a correct response (A_1E_1 or A_2E_2) on a trial when s_i is presented. Here we assume that the influence of each component of a new compound is proportional to that component's relative likelihood of eliciting a correct response on previous trials. For example, if s_1 has elicited more correct responses than s_2 on previous learning trials, then the response to the compound $(s_1 + s_2)$ will be more influenced by the s_1 cue than by s_2. The present hypothesis can be derived from an observing-response model (Atkinson, 1961) or from a perceptual-process model for discrimination learning (Atkinson, 1960). Again, as for Model I, we assume that with weight $1 - \omega$ there is a tendency to regress toward random responding in the presence of a new stimulus compound.

Model III

$$p_{ij} = \begin{cases} \omega p_i + (1-\omega)\tfrac{1}{2}, \text{ if } c_i > c_j \\ \omega p_j + (1-\omega)\tfrac{1}{2}, \text{ if } c_j > c_i \end{cases} \qquad [3]$$

$$p_{123} = \begin{cases} \omega p_1 + (1-\omega)\tfrac{1}{2}, \text{ if } c_1 > c_2 \text{ and } c_3 \\ \omega p_2 + (1-\omega)\tfrac{1}{2}, \text{ if } c_2 > c_1 \text{ and } c_3 \\ \omega p_3 + (1-\omega)\tfrac{1}{2}, \text{ if } c_3 > c_1 \text{ and } c_2 \end{cases}$$

(When equality holds in the relations among the c_i's, then apply Model I.) Here we assume that the response to a stimulus compound is determined solely by the single component cue that has been most frequently associated with a correct response on previous trials. For example, if s_1 has elicited more correct responses than s_2, then $p_{12} = p_1$. Again, as for both of the previous models, we assume that with weight $1 - \omega$ there is some regression toward a chance response level.

To illustrate predictions for each of these hypotheses consider a case in which $\pi_1 = 1.0$ and $\pi_2 = .50$. Further, assume that $\omega = 1$ and that after a large number of trials p_1 approaches a fixed value of 1, and p_2 a value of .50. Then $c_1 = 1$, $c_2 = .50$ and the predicted value for p_{12} will be $.50(1.50) = .75$ by Model I, $.66(1) + .33(.50) = .87$ by Model II, and 1 by Model III. It should be remarked that the only research we are aware of that utilizes an experimental methodology directly relevant to an evaluation of these models is that reported by Schoeffler (1954). However, in his study the particular parameter values employed generate identical predictions for all three models.

METHOD

Subjects and apparatus.—The *S*s were 48 children (26 girls and 22 boys) from the elementary school at the University of California, Los Angeles; all *S*s were in the fifth and sixth grade. They were randomly assigned to three groups with 16 per group.

The *S*s were run in subgroups of two or three with each *S* seated in a private booth. The apparatus, viewed from within *S*'s booth, consisted of a shelf at table level that was 30 in. wide and 10 in. deep. A panel 30 in. wide and 28 in. high was mounted vertically on the back edge of the shelf. Three signal lights (the s_i stimuli) were in a horizontal row, centered on the vertical panel; the lights were 15 in. from the base of the panel and were spaced at 2-in. intervals. Two silent operating keys (the A_1 and A_2 responses) were each mounted 2 in. in from the front edge of the shelf; these keys were 14 in. apart. One inch behind each of the keys was mounted a white frosted panel light (the E_1 and E_2 events).

Experimental design.—Each *S* was run for 960 trials. In consecutive blocks of 96 trials, there were 72 trials where s_1, s_2, and s_3 occurred equally often; on the remaining 24 trials the compounds $(s_1 + s_2)$, $(s_1 + s_3)$, $(s_2 + s_3)$, and $(s_1 + s_2 + s_3)$ occurred equally often. Thus, for example, in every 96-trial block s_1 was presented 24 times and $(s_1 + s_2)$ was presented 6 times. Otherwise, the presentation order of stimulus events was randomly determined for each *S*.

Three main groups were run. For all groups $\pi_1 = 1.0$ and $\pi_2 = .33$. The groups differed with respect to the value of π_3. For Group 1, $\pi_3 = .25$; for Group 2, $\pi_3 = .50$; and for Group 3, $\pi_3 = .75$. Within each of the main groups, two subgroups of eight *S*s were formed by counterbalancing the right and left positions of the response keys.

Procedure.—For each *S* one of the three signal lights was randomly designated s_1, another s_2, and the remaining one s_3. The *S*s were read the following instructions:

> This is an experiment to find out how good you are at guessing. It may be very much like other guessing games you've played before. Take a look at the board in front of you. One or more of the top lights go on every few seconds. Also note the two buttons with a light over each. When the experiment starts, and one or more of the top lights go on, you are to guess which of the bottom two lights will follow. You do this by pressing the button under the light you think will follow. If you think the left light will go on, press the left button; if you think the right light will go on, press the right button. Remember, as soon as one or more of the top lights go on, press the button on the side on which you think a light will go on, and see how many times you can guess correctly. Be sure to make your guess as soon as the top light or lights go on. If you are right a light will go on over the button you have pressed. If you were wrong, a light will go on over the other button. Try to make as many correct guesses as possible. Sometimes no light will appear over either button, but still you should have made a guess because you may have been correct. You will not know which one was correct but we are still keeping score. Now be sure to press the button as soon as the top light or lights go on and try to get as many correct as possible, but be sure to press one button every time. Are there any questions?

Questions were answered by paraphrasing the appropriate parts of the instructions. Following the instructions, 240 trials were run in continuous sequence. This was followed by a 5-min. rest period. After the rest, 240 additional trials were run. All *S*s were required to return the next day and another 480 trials were run with a 5-min. break between the first and last 240 trials. Thus each *S* was run a total of 960 trials.

On all trials the signal light (or lights) was lighted for 3 sec.; the time between successive signal onsets was 6 sec. The E_1 or E_2 light (if lighted) went on immediately following the offset of the signal light and remained on for 2 sec. The presentation of stimulus and reinforcing events and the recording of responses were automatically controlled.

RESULTS AND DISCUSSION

Mean learning curves and asymptotic results.—The top panel of Fig. 1 presents the proportion of A_1 re-

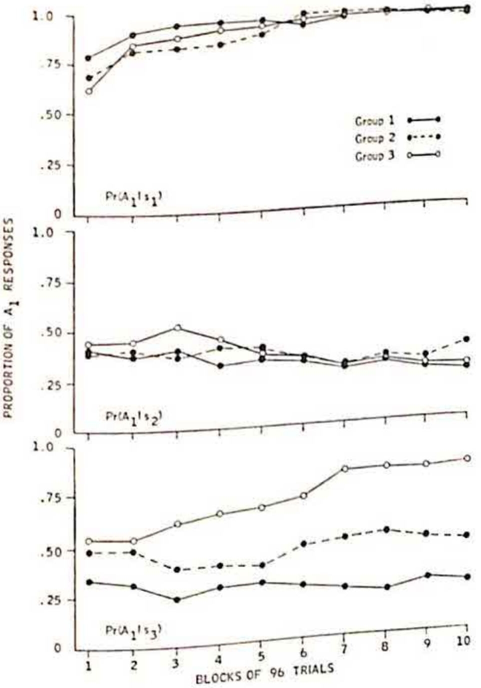


FIG. 1. Proportion of A_1 responses on s_i trials.

sponses on s_1 trials in successive 96-trial blocks. For each *S* the proportion of A_1 responses on s_1 trials was tabulated for a 96-trial block, and these quantities were averaged over *S*s. Similarly, the middle and bottom panels of Fig. 1 present the proportion of A_1 responses on s_2 and s_3 trials, respectively. In all three panels the curves appear to be reasonably stable over the last three or four blocks of trials; consequently proportions computed over the last 384 trials will be used as estimates of asymptotic performance.

Tables 1, 2, and 3 present the observed values of the probability of an A_1 response over roughly the last 384 trials to the individual stimuli and also to the stimulus compounds. Specifically what is presented is the proportion of A_1 responses to the individual stimuli for the last 100 presentations of each stimulus; similarly, the proportions for the stimulus compounds are based on the last 25 presentations of each stimulus compound.

For all groups the reinforcement schedules are the same for the s_1 and s_2 stimuli, and this similarity is reflected in the observed mean values of p_1 and p_2 given in Tables 1, 2, and 3. There are no significant differences

TABLE 1

OBSERVED PROPORTIONS FOR *S*s IN GROUP 1

S	p_1	p_2	p_3	p_{12}	p_{13}	p_{23}	p_{123}	m(I)	m(II)	m(III)	ω(I)	ω(II)	ω(III)
1	1.00	.52	.20	.84	.64	.32	.60	.047	.027	.084	1.000	.996	.280
2	1.00	.07	.23	.72	.80	0	.52	.152	.084	.148	1.000	1.000	.440
3	1.00	.16	.14	.60	.60	.28	.52	.042	.006	.057	.629	.628	.200
4	.99	.08	.04	.56	.56	.24	.48	.037	.009	.071	.591	.589	.122
5	.97	.47	.25	.72	.64	.40	.60	.027	.003	.036	1.000	.739	.298
6	.99	.45	.08	.68	.56	.40	.44	.047	.039	.072	.818	.546	.123
7	.97	.37	.52	.56	.52	.56	.72	.080	.076	.069	.082	.235	.128
8	.99	.52	.19	.84	.64	.32	.60	.051	.022	.083	1.000	1.000	.286
9	1.00	.26	.57	.68	.76	.44	.64	.030	.008	.037	.913	.716	.360
10	.98	.18	.67	.68	1.00	.40	.76	.113	.056	.098	1.000	1.000	.541
11	.97	.21	.08	1.00	1.00	.12	1.00	.373	.313	.033	1.000	1.000	1.000
12	.99	.46	.25	.72	.64	.40	.60	.025	.004	.037	.978	.690	.286
13	.97	.25	.03	.68	.24	.08	.20	.152	.169	.290	1.000	1.000	.001
14	.97	.15	.44	.52	.52	.48	.52	.008	.006	.001	.098	.093	.043
15	1.00	.11	.04	.72	.68	.08	.48	.105	.042	.124	.988	.983	.360
16	.96	.31	.27	.60	.56	.44	.56	.033	.017	.018	.521	.315	.131
M	.98	.29	.25	.69	.65	.31	.58	.082	.055	.079	.789	.721	.287

TABLE 2

OBSERVED PROPORTIONS FOR *S*s IN GROUP 2

S	p_1	p_2	p_3	p_{12}	p_{13}	p_{23}	p_{123}	$m(I)$	$m(II)$	$m(III)$	$\omega(I)$	$\omega(II)$	$\omega(III)$
1	.97	.34	.48	.68	.72	.44	.64	.025	.006	.020	.978	.746	.383
2	1.00	.43	.54	.80	.84	.92	.68	.153	.123	.156	1.000	.981	.600
3	1.00	.26	.41	.72	.80	.32	.88	.131	.062	.049	1.000	.989	.600
4	1.00	.53	.45	.40	.28	.24	.36	.180	.180	.180	.001	.001	.001
5	.99	.10	.95	.88	1.00	.60	.92	.170	.148	.141	1.000	1.000	.857
6	.97	.63	.49	.60	.56	.52	.60	.013	.014	.016	.333	.277	.213
7	.93	.34	.56	.92	.68	.52	.52	.110	.108	.122	.735	.599	.418
8	1.00	.34	.52	.68	.72	.48	.60	.019	.017	.039	.846	.647	.360
9	.99	.17	.09	1.00	.88	.36	.16	.286	.274	.239	.379	.383	.775
10	1.00	.01	.45	.60	1.00	.12	.56	.138	.069	.181	1.000	1.000	.200
11	1.00	.91	.75	1.00	.92	.80	.92	.038	.012	.043	1.000	1.000	.840
12	.92	.30	.39	.48	.72	.24	.60	.091	.077	.113	1.000	.942	.238
13	1.00	.35	.62	.16	.48	.60	.84	.200	.200	.200	.001	.001	.001
14	.99	.19	.43	.60	.72	.40	.56	.033	.018	.049	1.000	.534	.204
15	.99	.30	.22	.52	.52	.32	.44	.051	.057	.063	.191	.086	.041
16	1.00	.07	.48	.64	.72	.32	.56	.045	.017	.055	.916	.674	.280
M	.98	.33	.49	.67	.72	.45	.61	.105	.086	.104	.711	.616	.376

TABLE 3

OBSERVED PROPORTIONS FOR *S*s IN GROUP 3

S	p_1	p_2	p_3	p_{12}	p_{13}	p_{23}	p_{123}	$m(I)$	$m(II)$	$m(III)$	$\omega(I)$	$\omega(II)$	$\omega(III)$
1	1.00	.11	.89	.68	.92	.64	.76	.093	.052	.076	.944	.924	.520
2	1.00	.56	.86	.88	.84	.76	.68	.079	.070	.054	.791	.767	.680
3	1.00	.08	.97	.72	.88	.40	.76	.106	.081	.126	.784	.780	.520
4	.99	.30	.97	.64	.76	.60	.68	.031	.005	.048	.710	.561	.367
5	.99	.30	.42	.44	.36	.36	.36	.120	.120	.120	.001	.001	.001
6	1.00	.03	.97	.60	.84	.36	.92	.133	.110	.195	.701	.698	.680
7	1.00	.14	.95	.56	.72	.56	.60	.019	.015	.048	.463	.459	.200
8	1.00	.14	.95	.60	.72	.76	.60	.079	.066	.073	.463	.459	.200
9	1.00	.52	.68	.84	.88	.56	.72	.043	.030	.056	1.000	.996	.680
10	.97	.60	.96	.80	.88	.84	.72	.060	.047	.052	.817	.816	.639
11	.94	.38	.69	.76	.52	.56	.72	.118	.091	.069	1.000	.988	.500
12	.96	.37	.89	.84	.84	.84	.24	.224	.212	.163	.800	.790	.739
13	1.00	.42	.49	.56	.56	.48	.52	.008	.010	.013	.245	.181	.120
14	.96	.24	.75	.56	.60	.52	.52	.019	.016	.023	.282	.265	.130
15	.98	.15	.94	.76	.58	.28	.52	.123	.122	.133	.174	.173	.166
16	.99	.39	.91	.72	.80	.56	.72	.044	.027	.051	.667	.657	.449
M	.99	.30	.83	.68	.73	.57	.63	.081	.067	.081	.615	.595	.412

among the three groups on the p_1 measure, $F\ (2, 45) = .05$, or the p_2 measure ($F = .24$). In contrast, the differences among the three groups on the p_3 measure are highly significant ($F = 37.72$) reflecting the effect of the π_3 variable.

Theoretical analyses.—We now turn to the problem of predicting the compound probabilities p_{12}, p_{13}, p_{23}, and p_{123} in terms of p_1, p_2, and p_3. As indicated earlier, we assume (for all three models) that there may be a regression effect associated with the introduction of a stimulus compound. This regression effect is defined by the parameter ω, and we must estimate ω separately under the assumptions of each model. The estimate of ω was made for individual *S*s by a method that is equivalent to a least squares procedure. Specifically, let $p_{ij}(\omega; \mathrm{I})$ be the prediction for stimulus compound $(s_i + s_j)$ generated by Model I as a function of ω; e.g.,

$$p_{12}(\omega; \mathrm{I}) = \omega[\tfrac{1}{2}(\hat{p}_1 + \hat{p}_2)] + (1 - \omega)\tfrac{1}{2},$$

where $\hat{p}_1$ and $\hat{p}_2$ denote the observed proportions for a particular *S*. Then for each *S* we define the function

$$\begin{aligned} m(\omega; \mathrm{I}) = \tfrac{1}{4}\{ &|p_{12}(\omega; \mathrm{I}) - \hat{p}_{12}| \\ &+ |p_{12}(\omega; \mathrm{I}) - \hat{p}_{13}| \\ &+ |p_{23}(\omega; \mathrm{I}) - \hat{p}_{23}| \\ &\quad + |p_{123}(\omega; \mathrm{I}) - \hat{p}_{123}|\}, \end{aligned}$$

where the $\hat{p}_{ij}$'s are the observed proportions on the stimulus compounds. The quantity $m(\omega; \mathrm{I})$ gives the average absolute difference between the predicted and observed values as a function of ω. For each *S* an ω was selected that minimized the function $m(\omega; \mathrm{I})$; this was done by computing the function for $\omega = .001k$ $(k = 0\text{–}1{,}000)$ on a high-speed digital computer and selecting the ω that gave the minimum value. The minimum value of the function is denoted $m(\mathrm{I})$ and the value of ω associated with the minimum is called $\omega(\mathrm{I})$. Using precisely the same methods, we obtain $m(\mathrm{II})$, $\omega(\mathrm{II})$, $m(\mathrm{III})$, and $\omega(\mathrm{III})$ for each *S* under the assumptions of Models II and III, respectively.

Individual estimates of ω and the corresponding minimum value of $m(\cdot)$ are given in Tables 1, 2, and 3. For a particular *S* one can select the model that provides the best fit; i.e., the model that gives the smallest average absolute deviation. For example, the first *S* in Group 1 has the values $m(\mathrm{I}) = .04$, $m(\mathrm{II}) = .03$, and $m(\mathrm{III}) = .08$; for this *S* Model II gives the best fit, Model I is next best, and Model III provides the poorest fit. If the information presented in Tables 1, 2, and 3 is tabulated in this fashion, then for Group 1 the best fit is given by Model II for 12 *S*s, by Model III for 3 *S*s, and by Model I for 1 *S*. For Group 2, the best fit is given by Model II for 9 *S*s, Model III for 3 *S*s, and Model I for 4 *S*s. For Group 3, Model II is best for 11 *S*s, Model III for 3 *S*s, and Model I for 2 *S*s. Combining results for all three groups, Model II gives the best fit for 32 *S*s, Model III for 9, and Model I for 7.

TABLE 4

PAIRWISE COMPARISONS OF THE THREE MODELS ON THE MINIMUM OF $m(\cdot)$

Group	Models I and II		Models I and III		Models II and III	
	I	II	I	III	II	III
1	1	15	10	6	13	3
2	4	12	12	4	13	3
3	2	14	10	6	13	3
Total	7	41	32	16	39	9

TABLE 5

OBSERVED AND PREDICTED PROPORTIONS FOR MODELS I, II, AND III

Group	Model	p_{12}	p_{13}	p_{23}	p_{123}
1	I	.61	.58	.31	.50
	II	.66	.63	.31	.55
	III	.64	.64	.40	.64
	Ob	.69	.65	.31	.58
2	I	.60	.68	.44	.57
	II	.64	.70	.44	.61
	III	.68	.68	.43	.68
	Ob	.67	.72	.45	.61
3	I	.59	.76	.56	.64
	II	.64	.76	.57	.67
	III	.70	.70	.65	.70
	Ob	.68	.73	.57	.63

Possibly a better comparison is provided by considering the three models pairwise on $m(\cdot)$ for individual *S*s; i.e., comparing Models I and II, then Models II and III, and finally Models I and III. Again these comparisons can be made by inspecting Tables 1, 2, and 3. Table 4 presents the results of such an analysis. As we see, Model II is better than Model I for 41 of the 48 *S*s and Model II is better than Model III for 32 of the 48 *S*s. In terms of these tabulations Model II is clearly the best predictor among the three models.

Table 5 presents the average of the observed p_{ij}'s and also the theoretical values for the three models. Predicted values for each *S* were computed (using the estimates of ω given in Tables 1–3) and the averages of these values are presented in Table 5. An inspection of Table 5 suggests that Model II provides the best account of our data, but even for this model some sizable discrepancies exist between observed and predicted probabilities. To obtain a direct measure of the overall goodness of fit one can puruse the values of $m(\mathrm{I})$, $m(\mathrm{II})$, and $m(\mathrm{III})$ given in Tables 1–3. An average over all 48 *S*s yields the following values: $\overline{m(\mathrm{I})} = .089$, $\overline{m(\mathrm{II})} = .069$, and $\overline{m(\mathrm{III})} = .088$. Here again, the statistic favors Model II and indicates that the absolute difference between predicted and observed quantities over all *S*s was on the average about .07.

In evaluating the fit of Model II, we also are interested in the obtained estimates of ω. In theorizing about any of the models examined in this paper, one would prefer to assume that ω is determined solely by the characteristics of the component stimuli plus background stimulation associated with the particular experimental situation, and is independent of other factors. This statement does not mean that we postulate no individual differences in ω over *S*s, but only that variations in ω should be independent of reinforcement schedules and stimulus-presentation schedules. This independence assumption and its implications for behavior theory are discussed in considerable detail elsewhere (see the section on stimulus compounding in Atkinson & Estes, 1963). For the present study, the above comments require us to predict no difference in the values of ω over the three experimental groups. By inspecting Tables 1–3 we note, for Model II, that the mean values of ω were .721, .616, and .595 for Groups 1, 2, and 3, respectively. An *F* test on these three groups of ω estimates yields a ratio of .68 which does not approach significance with $df = 2/45$.

In terms of the above considerations it appears that Model II provides the best fit to our data and that the regression effect characterized by the parameter ω is relatively constant over different reinforcement schedules. Unfortunately, even for Model II there are some fairly large discrepancies between theory and observation. Thus, it seems that other models for stimulus compounding need to be examined and it is for this reason that we have presented our observed results for individual *S*s. It is our hope that these data may prove useful in new models.

REFERENCES

Atkinson, R. C. A theory of stimulus discrimination learning. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences.* California: Stanford Univer. Press, 1960. Pp. 221–241.

Atkinson, R. C. The observing response in discrimination learning. *J. exp. Psychol.*, 1961, **62,** 253–262.

Atkinson, R. C., & Estes, W. K. Stimulus sampling theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology.* Vol. 2. New York: Wiley, 1963. Pp. 121–268.

Burke, C. J., & Estes, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22,** 133–145.

Schoeffler, M. S. Probability of response to compounds of discriminated stimuli. *J. exp. Psychol.*, 1954, **48,** 323–329.

(Received February 1, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 59–66

TASTE THRESHOLDS, DETECTION MODELS, AND DISPARATE RESULTS[1]

EUGENE LINKER,[2] MARY E. MOORE
University of Pennsylvania

AND EUGENE GALANTER
University of Washington

We studied the effect on false alarms and correct detections in taste threshold determinations caused by variations in signal presentation probabilities. A modified method of constant stimuli was used. Results indicate that a simple threshold theory with correction for guessing cannot account for these data. A more sophisticated threshold theory can do as well as signal detectability theory.

Psychophysical measures of the taste threshold often yield discrepant results. Pfaffman (1951) has suggested that these discrepancies may be attributable to a number of factors, including the techniques used for applying the stimulus, the procedures for arranging the contingencies between the stimulating conditions and responses. Other factors that affect these results include sensory adaptation (Hahn, 1934), and altered states of body chemistry (Goetzl, Ahokas, & Payne, 1950; Hammer, 1951; Mayer-Gross & Walker, 1946; Yensen, 1959). Among the specific stimulus variables found to affect the threshold are area and duration of stimulation (Bujas, 1934; Bujas & Ostojcic, 1941) and the temperature of the solution (Hahn, 1936).

In the course of studies on threshold determination many methods for applying the stimuli have been devised. These range from techniques that affect the entire surface of the tongue to those that stimulate specific areas. Woodworth and Schlosberg (1954) report that methods of stimulation that involve large areas of the tongue yield the lowest thresholds. Because an absolute threshold is the bottom of a sensory continuum, methods that yield low threshold values are preferred. The sipping procedure is simple and also gives low absolute thresholds; it is a widely used technique.

In Table 1 we show a summary of the methods and their results for a sample of previous behavioral studies that have attempted to determine the absolute sucrose threshold for human *S*s.

These various methods yield thresholds that differ by as much as a factor of 3, and therefore it would be surprising if the experimental results reported in one laboratory could be repeated in another unless exactly the same procedure were used. In addition, from the published reports it is difficult to tell whether results can best be described in terms of detection or recognition thresholds. It appears that Fabian and Blum (1943), Richter and Campbell (1940), and Schutz and Pilgrim (1957) have concerned them-

---

[1] This research was supported in part by Grant M-3684, National Institute of Mental Health, National Institutes of Health, Public Health Service, and in part by the Office of Naval Research, Contract NONR-551(37). Reproduction in whole or part is permitted for any purpose of the United States Government.

[2] Now at Rutgers University.

TABLE 1

SUMMARY OF SELECTED METHODS AND RESULTS FROM STUDIES OF THE ABSOLUTE SUCROSE THRESHOLD FOR PEOPLE

Investigator	Stimulus Presented by	Method	Average Threshold	
			Mols × 10^3	% Weight/Volume
Richter and Campbell (1940)	Ad lib. sips	Method of limits. Paired solutions of sucrose and distilled water. Ascend to point where *S* reports "different." Then descend in smaller steps to lowest point *S* still reports "different." No catch trials	4.97	0.17
Fabian and Blum (1943)	5-cc sip	Method of limits. Two solutions (one distilled water). Two ascending series. Average of first trials on which *S* said "different." No catch trials	16.1 (geometric *M*)	0.55
Janowitz and Grossman (1949)	Ad lib. sips	Method of constant stimuli. If pair judged "different" *S* was instructed to name taste perceived	14.9 (before lunch)	0.51
Schutz and Pilgrim (1957)	6-cc sip	Method of constant stimuli	10.2	0.35
Yensen (1959)	8–10 cc sip	Method of limits. Four ascending series. One solution. *S* reported the presence or absence of sucrose. Two successive positive reports. Threshold halfway between lowest positive report and the next lowest concentration	8.2 (approximate for 11:30 A.M. testing)	0.28
Furchtgott and Friedman (1960)	1-cc sip	Method of limits. Three ascending series. Forced choice of sucrose from one of four solutions (3 tap or distilled water). Two correct identifications on successive solutions. No catch trials	8.2 (before lunch)	0.28

selves with detection thresholds. It is somewhat difficult to tell how to classify Yensen's (1959) results. In his experiment *S*s were told to respond in terms of the presence or absence of the taste quality under consideration (e.g., sucrose) and the stimuli consisted simply of a graded series of sucrose solutions. Because any deviation from a neutral taste would probably be interpreted by *S* as sucrose, the results are more reasonably characterized as detection rather than recognition. Furchtgott and Friedman's (1960, p. 577) study is also difficult to categorize. Their report states that *S* was presented with four beakers, three of which were either tap or distilled water and the fourth either sucrose, hydrochloric acid, or salt. "Progressively more concentrated solutions were presented until correct identification was achieved at two successive levels." Now, it is not altogether clear whether *S* was or was not told of the taste under consideration. If he was, the results should be considered as detection thresholds. If he was not, in view of possible expectancies established in the first series of a test session, succeeding series may have been some conglomeration of recognition and detection measurements. In the study by Janowitz and Grossman (1949), although the results are reported as recognition thresholds the same possibility of the confounding effects of "expectancy" exists. This was because *S* had more than one test session per day with the same taste quality.

This problem of consistency is more complex than the mere ascription of differences of results to differences of procedure and definition. The whole notion of the absolute threshold has been critically attacked by theoreticians concerned with the abstract detection problem itself (Tanner & Swets, 1954). Galanter (1962) has pointed out that in addition to the stimuli there are at least *S*'s expectations and motives that enter into the decision as to which of a set of alternative responses he will choose in a psychophysical threshold experiment. The purpose of the present experiment was to discover whether at least one of these factors, expectations, would indeed change the observed frequency of reporting for fixed sucrose concentrations in ways that sophisticated theories of detection would suggest.

METHOD

Subjects.—The *Ss* were three male graduate students in the philosophy department at the University of Pennsylvania. The experiment consisted of two parts—isobias determinations and isosensitivity determinations. Both parts used a straightforward signal detection procedure with a symmetrical payoff matrix.

From *S*'s point of view the task consisted of discriminating between a 1-cc sip of an aqueous sucrose solution (S) and a 1-cc sip of distilled water (W). For each correct response of "sugar" (s) or "water" (w), *S* received 5¢ and for each incorrect response 5¢ was subtracted from his winnings. Thus, *S* could win money by making a *hit*; i.e., reporting sugar when sugar was actually present—symbolized as (s/S) or by correctly identifying a sample of distilled water (w/W). The *S* could lose money by making a *false alarm*; i.e., reporting sugar when the stimulus was distilled water—symbolized as (s/W) or by reporting water when the stimulus was the sugar solution (w/S). The *Ss* were compensated for their services solely in terms of their winnings; for performances below chance where their winnings were actually negative amounts, they were paid nothing.

Daily test sessions which lasted about 50 min. consisted of 70 experimental trials. The *Ss* were blindfolded and wore earphones. They were seated alongside a table with their right arm positioned by a jig with a well that held a 30-cc "shot" glass. One cubic centimeter of the stimulus was delivered by a manually operated syringe into the "shot" glass. An auditory signal announced the beginning of a trial. The *S* picked up and drained the glass and then responded "sugar" or "water." He was provided with immediate feedback concerning the correctness of his response by another auditory signal consisting of a single "beep" for correct and a double "beep" for incorrect. Thirty seconds from the time of his response the signal for the next trial was given. This 30-sec. intertrial interval was used to reduce adaptation effects.

Test solutions were prepared from reagent grade sucrose and double distilled water by diluting a standard 10% solution which was prepared fresh daily. Temperatures of the stimuli were equated at the beginning of a test session at 75° F. and monitored during the course of the session. Maximum pre-post-temperature differences were ±2° F. However, the average differences were much smaller, and sessions in which the stimuli differed from each other by ≥1° F. were discarded.

Testing began for *S*1 and *S*2 at 11:00 A.M. and for *S*3 at 12:00 noon. The *Ss* were

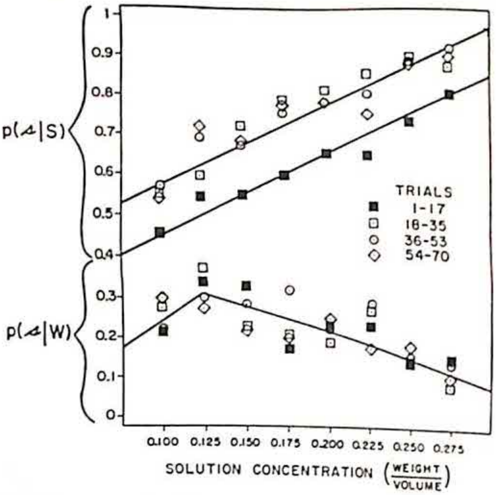


FIG. 1. $p(s/S)$ and $p(s/W)$ by quarters of experimental trials as a function of solution concentrations. (All points are the average of 3 days testing for each of three *S*s at each concentration. $P(S) \cong P(W) \cong .5$.)

asked to fast and to refrain from smoking for 11 hr. prior to the test session.

Isobias determinations.—Each of the *S*s served in 24 daily sessions after 6 days experience with the procedure. Each daily session involved testing with only one solution concentration. The *S*s were run on a random schedule for 3 days (210 trials) at each of eight solution concentrations which ranged from .100% (.00292M) to .275% (.0804M) in step intervals of .025%. They did not know the schedule, and they received the same instructions each day. They were always told that $P(S) \cong P(W) \cong .5$. Daily presentation schedules were derived from random number tables with the sole restriction that $P(S)$ varied between .41 and .59.

Isosensitivity determinations.—Isosensitivity curves were generated by varying the presentation probabilities for distilled water $[P(W)]$ and for a selected sucrose concentration $[P(S)]$. In order to sweep out the curves sucrose concentrations were selected for each *S* which had produced $p(s/S) \cong .75$ and $p(s/W) \cong .25$ for $P(S) \cong P(W) \cong .5$. For *S*1 and *S*2 this was a sucrose concentration of .200% (.00584M) and for *S*3 a concentration of .225% (.00658M). The *S*s were informed of the approximate value of $P(S)$ and $P(W)$ at each session prior to the actual testing. The schedules for $P(S)$ and $P(W)$, respectively, were .1–.9, .3–.7, .5–.5, .7–.3, .9–.1. Sequential orders of S and W were derived from tables of random numbers with the sole restriction that the empirical probabilities did not differ from the stated ones by more than approximately ±.05.

For *S*1, 280 trials were given for Schedules .1–.9 and .9–.1, 560 for .5–.5, and 350 for .3–.7 and .7–.3. The data for the .5–.5 presentation schedule includes the 3 days of testing used for the isobias determinations. The extra testing at Schedules .5–.5, .3–.7, and .7–.3 was part of another experiment being conducted simultaneously in which this procedure was identical.

The procedure for *S*s 2 and 3 was somewhat different. Because the presence of a "warm-up" effect for taste was found in the isobias data, a warm-up period was provided for *S*s 2 and 3 prior to each daily session. This consisted of five 1-cc samples of the sucrose solution and five samples of the distilled water presented alternately every 30 sec. and identified for *S* prior to his tasting them. As for *S*1, a randomly determined test order for the experimental presentation schedules was used and four test sessions (280 trials) were conducted at each of the five schedules.

RESULTS

We consider first certain questions about the stability of the results. Our initial results confirmed the Pfaffman, Schlosberg, and Cornsweet (1954) finding of a warm-up effect. This effect is shown in Fig. 1. Consider first the upper four curves. The filled squares are the average $p(s/S)$ over *S*s for the first 17 trials of the three experimental sessions. The other upper curves are for subsequent quarters. Inspection shows a consistently lower $p(s/S)$ for the first quarter. The four lower curves are similar averages for values of $p(s/W)$. Surprisingly, no warm-up effect is obvious for these false alarms. We offer no explanation for the appearance of a warm-up effect on sucrose trials and its failure to appear on water trials. We are able to assert, however, that the warm-up effect is a change in sensitivity rather than a change in *S*'s criterion. A subsequent attempt to eliminate this warm-up effect by giving $S2$ and $S3$ 10 practice trials did not prove particularly effective.

We can also ask whether there is a shift among sessions in the response probabilities. Figure 2 shows no such change.

The absolute threshold estimated from psychometric functions is ca. .200% sucrose, which agrees fairly well with the Richter and Campbell (1940) results using ad lib. sips with the method of limits. The agreement is probably even better for on approximately 20% of the trials, at the threshold concentration, S was reporting sugar even though water alone was presented. Richter and Campbell's lower value may be attributable to the fact that S was reporting the presence of sugar when in fact he did not taste it. The S's false alarms decrease as the sugar concentration increases just as signal detectability theory would suggest. That is to say, S appears to show a consistent bias towards saying "yes" with a fixed probability that is contingent upon the sucrose concentration. We plot these data in Fig. 3 to reveal this "isobias" function. For varying concentrations we see that the points move more or less along the 45° line. This 45° line, a line of constant response bias, is one of a family of curves that we could draw in this figure, the form of which would be contingent upon the particular theory that we use. Because of the small number of observations at each point, it seems unnecessary to attempt a better fit.

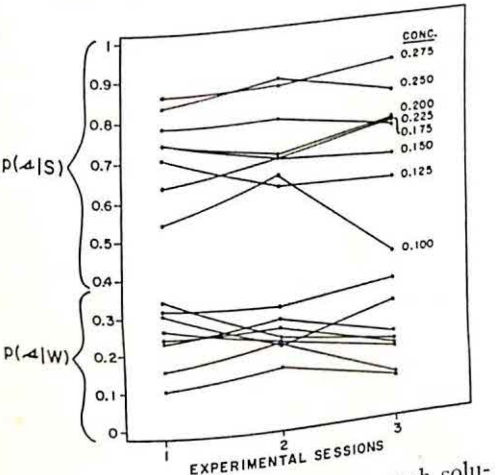


FIG. 2. $p(s/S)$ and $p(s/W)$ at each solution concentration averaged over Ss for each of the three experimental sessions.

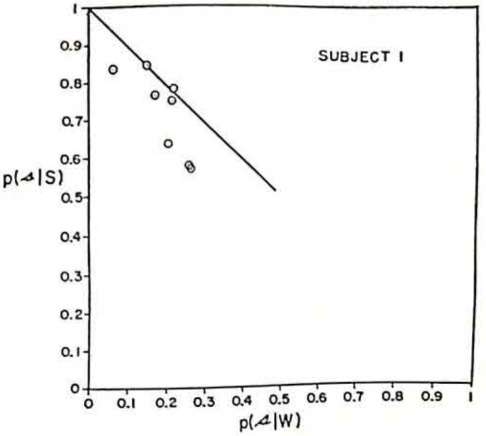


FIG. 3. $p(s/S)$ as a function of $p(s/W)$ at each solution concentration for $S1$. ($P(S) \cong P(W) \cong .5$.)

Isosensitivity determinations.—The results as far as we have taken them show that with some reliability S correctly reports the presence of sugar with increasing probability as the concentration of sugar increases, just as the classical view of things would suggest. In addition, however, S does make false alarms and these false alarms behave in a relatively systematic manner depending upon the concentration of sugar presented. These results, of course, are those that were obtained with a presentation probability of .5. We now ask the question, for a fixed concentration of sugar, can we vary the sucrose detection probability and the false alarm rate by varying the signal presentation probability? That is to

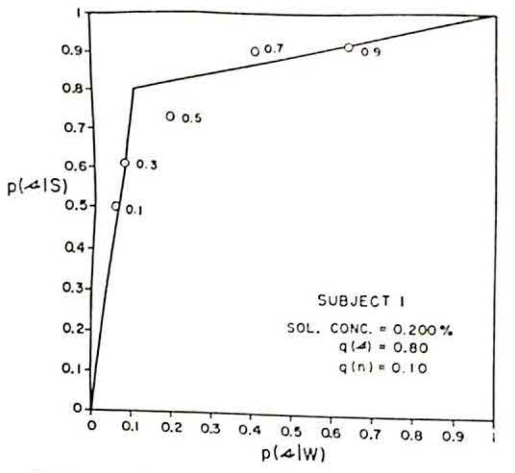


FIG. 4. The isosensitivity function for *S*1. (Each point is labeled with P(S). The straight lines represent theoretical functions derived from Luce's threshold model.)

say, can we alter *S*'s expectations of sugar and thereby alter his probability of responding that sugar is present?

For these isosensitivity determinations, we varied the signal presentation probability over five values using a single concentration of sucrose, the concentration for which p(s/S) was approximately .75 and p(s/W) was approximately .25 during the isobias determinations. Figures 4 and 5 are the isosensitivity functions for two of the *S*s. These points are labeled by P(S). The straight lines are theoretical functions to be described subsequently.

Looking at Fig. 5 as an example, we see that for a single sucrose concentration of .225%, the probability of reporting the presence of sugar varies from about .26 to about .96. The high probability of report is associated with a high presentation probability. Thus, if as in classical experiments this sucrose concentration had been continuously presented, *E* might well have thought that it was very easily detectable. On the other hand, the fact that at this concentration and a signal presentation probability of .9, approximately 77% of the time *S* reported the presence of sugar when none at all was present casts serious doubt on the reasonableness of this interpretation. Thus we see that for a fixed sucrose concentration the ostensible threshold varies over a tremendous range. We can, of course, select some arbitrary presentation probability as the "standard" and then vary the concentrations as we did for the isobias determinations and use some arbitrary cut-point as a representation of the threshold. It seems more reasonable, however, to investigate the whole decision process itself within the context of the psychophysical design if we hope to explicate the role of the sensory processes.

DISCUSSION

In this, as in all psychophysical experiments (although often implicit), there were three interlocking independent variables. Only two of them, signal strength (sucrose concentration) and the signal presentation probability, were varied; the outcome structure in this experiment was held fixed. Thus, whatever effect the outcomes may have had upon the behavior, that effect was presumed to be

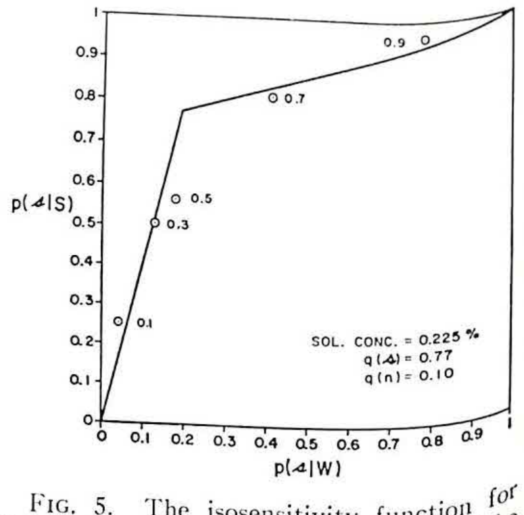


FIG. 5. The isosensitivity function for *S*3. (Each point is labeled with P(S). The straight lines represent theoretical functions derived from Luce's threshold model.)

equivalent for all variations in the experiment.

On the response side of the experiment, there were two independent classes of responses. These were correct and incorrect responses associated with the presence of sucrose and correct and incorrect responses associated with its absence. These responses are construed as representing two kinds of theoretical processes. On the one hand, we assume that from time to time S makes a response that faithfully reflects the state of his interior vis-a-vis the sensory effects of the stimulation. On the other hand, certain of the responses can be construed as reflecting the action of events more appropriately assigned to central cognitive processes, and therefore only effective upon those variables that modulate these hypothecated events. If we are concerned with the effects of stimulation, then we must conjecture some theory that will enable us to decompose the behavior into that part of it that is properly contingent upon the sensory effects themselves, and that part contingent upon other things.

Classically, E often assumed that all the responses were dependent only on the sensory events, and that in those cases where he found evidence of a failure of this dependence he assigned it to inadequate training of S. This experiment demonstrates that for the taste modality, as others have shown for other modalities, such an interpretation is untenable. When this untenability has been suspected in the past, a simple theory has been proposed to extract out the "spurious" behavior from the truly sensory. This theory was often called a "correction for guessing." The assumptions that underly it assert that the true probability of saying "sugar" when no sugar is present is 0, but that S makes false alarms. Therefore we can correct the observed proportion of hits to get the theoretical detection probability by taking the difference between the proportion of hits and false alarms and dividing that difference by $1 - p(\mathrm{s}/\mathrm{W})$. However, the consequence of such a theory is that the form of the isosensitivity curve would be a straight line running from some point on the ordinate (specifically the true probability of reporting the signal when it is present) to the upper right hand corner of the square. Figures 4 and 5 show that such a correction for guessing is not appropriate for these experiments.

The results that we obtain are more consistent with the general theory of signal detectability (cf. Tanner & Swets, 1954). Luce (1963), however, suggests a theory that, like correction for guessing, presumes the existence of a threshold (a physiologically and theoretically viable idea) but that assumes the introduction of bias on S's responses in ways that generate the kinds of behavior we observe. The theory, in abbreviated form, is as follows: We assume that when a signal is presented on a given trial there exists a true probability $q(s)$ that S will detect the presence of the signal. It is further assumed that S introduces two mutually exclusive response biases. He can falsely convert a fixed proportion of his detection observations into negative responses or some failures of detection into positive responses. There is a true detection probability $q(n)$ when S detects a signal when none is presented. The consequence of this theory is an isosensitivity curve that consists of two line segments, one running from the point $(0, 0)$ to $[q(n), q(s)]$ that Luce calls the lower limb, and the upper limb from $[q(n), q(s)]$ to $(1, 1)$. These lines are drawn by eye for the data of the present experiment and the values $q(n)$ and $q(s)$ that constitute the parameters of the curve are shown on the figures. It should be noted that this theory divides the classical correction for guessing threshold theory into two components corresponding to the two true detection probabilities.

The consequences of this experiment are twofold. On the one hand, they reveal that classical methods for estimating psychophysical thresholds are inexorably confounded with biases on S's responses that may be independent of the effects of the stimulus upon the sensory system. The S's responses bear orderly, but not necessarily direct, relations to the stimulating conditions. As a consequence, the use of psychophysical data as a baseline for establishing physiological parameters is unjustifiable without assuming some underlying psychophysical theory to account for the effects of these nonsensory events on the responses that *S*s generate in the psychophysical setting. A second consequence of this experiment is to show factually that the simple threshold model which has been proved inadequate in other sensory modalities like hearing (Tanner, Swets, & Green, 1956) or vision (Swets, Tanner, & Birdsall, 1955, 1961) is also inappropriate in the taste mode.

REFERENCES

BUJAS, Z. Le temps d'action des stimuli de la sensibilité gustative. *CR soc. Biol., Paris*, 1934, **117**, 1307–1309.

BUJAS, Z., & OSTOJCIC, A. La sensibilité gustative en function de la surface excité. *Acta Inst. Psychol. U. Zagreb.*, 1941, **13**, 1–19.

FABIAN, F. W., & BLUM, H. D. Relative taste potency of some basic food constituents and their competitive and compensatory action. *Food Res.*, 1943, **8**, 179–193.

FURCHTGOTT, E., & FRIEDMAN, J. The effects of hunger on taste and odor RLs. *J. comp. physiol. Psychol.*, 1960, **53**, 576–581.

GALANTER, E. Contemporary psychophysics. In, *New directions in psychology*. New York: Holt, Rinehart, & Winston, 1962.

GOETZL, F. R., AHOKAS, A. J., & PAYNE, J. G. Occurrence in normal individuals of diurnal variations in acuity of the sense of taste for sucrose. *J. appl. Physiol.*, 1950, **2**, 619–625.

HAHN, H. Die Adaptation des Geschmacksinnes. *Z. Sinnesphysiol.*, 1934, **65**, 105–145.

HAHN, H. Über die Ursache der Geschmacksempfindung. *Klin. Wschr.*, 1936, **15**, 933–935.

HAMMER, F. J. The relation of odor, taste, and flicker-fusion thresholds to food intake. *J. comp. physiol. Psychol.*, 1951, **44**, 403–411.

JANOWITZ, H., & GROSSMAN, J. Gusto-olfactory thresholds in relation to appetite and hunger sensation. *J. appl. Physiol.*, 1949, **2**, 217–222.

LUCE, R. D. A threshold theory for simple detection and recognition experiments. *Psychol. Rev.*, 1963, **70**, 61–79.

MAYER-GROSS, W., & WALKER, J. W. Taste and taste selection of food in hypoglycemia. *Brit. J. exp. Pathol.*, 1946, **27**, 297–305.

PFAFFMAN, C. Taste and smell. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1143–1171.

PFAFFMAN, C., SCHLOSBERG, H., & CORNSWEET, J. Variables affecting difference tests. *Food Accept. test. Methodol.*, 1954 (Oct.), 4–17.

RICHTER, C. P., & CAMPBELL, K. Sucrose taste thresholds of rats and humans. *Amer. J. Physiol.*, 1940, **128**, 291–297.

SCHUTZ, H. G., & PILGRIM, F. J. Differential sensitivity in gustation. *J. exp. Psychol.*, 1957, **54**, 41–48.

SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. The evidence for a decision-making theory of visual detection. Technical Report No. 40, 1955, University of Michigan, Electronic Defense Group.

SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 301–340.

TANNER, W. P., JR., & SWETS, J. A. A decision making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

TANNER, W. P., JR., SWETS, J. A., & GREEN, D. M. Some general properties of the hearing mechanism. Technical Report No. 30, 1956, University of Michigan, Electronic Defense Group.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.

YENSEN, R. Some factors affecting taste sensitivity in man: I. Food intake and time of day. *Quart. J. exp. Psychol.*, 1959, **11**, 221–229.

(Received February 4, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 67–71

MEDIATION OF R_1-R_2 IN THE ACQUISITION OF S-R_1, S-R_2 ASSOCIATIONS[1]

NANCY MORNING AND JAMES F. VOSS[2]

College of Wooster

The development of B-C associative strength was studied in Stage 2 following Stage 1 training in which 6 A-B, A-C modified paired associates, each with a .5–.5 probability ratio, were presented for 0, 40, or 80 trials. In Stage 2, 2 paired associates were B-C, 2 were B-D (control), and 2 were B-Cr where the C items were presented with different Stage 1 stimuli. The results indicate that positive mediational effects increased as Stage 1 training increased, with B-Cr yielding significantly more correct responses than B-Cr following 40 Stage 1 trials and more than both B-Cr and B-D with 80 Stage 1 trials. Negative mediational effects were significant and greater after 40 pretraining trials than after 80. The results were interpreted in terms of Postman's (1961) retroaction-transfer theory.

Voss, Thompson, and Keegan (1959) and Goss and Sugarman (1961), using a modified paired-associate paradigm in which a stimulus word was presented on each trial with one of two response words, found that performance varied as a function of the probability ratio of the two response words. Horton and Kjeldergaard (1961), using a three-stage A-B, A-C (experimental) or D-C (control), B-C mediation paradigm, found significantly better B-C performance following A-C rather than D-C. Since the Voss et al. design of virtually simultaneous A-B, A-C acquisition was similar to the successive A-B, A-C acquisition of Horton and Kjeldergaard (1961), it was hypothesized that B-C acquisition is facilitated by simultaneous A-B, A-C training.

Norcross and Spiker (1958), using a three-stage A-B, B-C, A-Cr (experimental) or A-D (control) paradigm, in which Cr represents C items paired with A items where B was not common to A and C, found poorer performance for A-Cr than A-D. Considering the Voss et al. simultaneous A-B, A-C paradigm, it was hypothesized that negative mediational effects occur when the B-C pairing involves B and C items which were previously associated with different stimulus items (A).

Two factors were considered necessary to demonstrate the hypothesized mediational effects: first, training on A-B, A-C is required; second, performance on B-C is different in the hypothesized direction from performance on B-D, where D is a control item.

METHOD

Procedure.—Three groups, each of $N = 30$, received 0 (Group 0), 40 (Group 40), or 80 (Group 80) trials of Stage 1 training, respectively. In Stage 1, a modified paired-associate procedure was employed in which a stimulus, A, was followed by A and one of two responses (A-B or A-C). On each trial, six such associations were presented in random order. The probability ratio of B and C was .5–.5 for each of the six associations. The six sequences of B and C were each randomized within 10-trial blocks for 40 trials. The 80-trial sequence constituted two presentations of the 40-trial list. The list was presented via strip film at a 2-sec. rate with a 2-sec. intertrial interval.

---

[1] This experiment was supported by National Institute of Mental Health Research Grant M-3531.

[2] Now at the University of Pittsburgh.

The instructions of Stage 1 were:

> Your task in this phase of the experiment will be to learn to anticipate which word will be paired with a given word. The procedure is as follows: a word will be presented, e.g., MAN. This word will be followed by one of two words, e.g., BUSH or STREET, paired with the first word. Your task is *to learn to anticipate* which of the two words will occur with the first word. The two words occur equally often, so it is important that you use both words in order to obtain a large number of correct anticipations. More than one such association will be presented to you.
>
> Please note that you should respond as often as possible by indicating which word will be paired with the first word and be sure to use both response words associated with each stimulus word.
>
> Are there any questions?

Following Stage 1 completion, *S*s were conducted to a second room and instructed to sort blocks on the Minnesota Spatial Relations Test. The intervening task time was controlled so that *S*s returned to the original experimental room and began Stage 2 5 min. after the completion of Stage 1.

Stage 2 involved presentation of six paired associates for 20 trials. Five random trial orders of the six paired associates were employed at a 2-sec. rate with a 2-sec. intertrial interval.

Two of the six paired associates of Stage 2 were the B-C items of a particular A-B, A-C association of Stage 1 (B-C); two were the B and C items of Stage 1, but the Stage 2 pairings were such that the B and C items were presented with different stimuli (A) in Stage 1 (B-Cr); the remaining two paired associates were B-D, where D was a word not previously used.

The instructions for Stage 2 were:

> This phase of the experiment is similar to the first phase except that only one word will occur with the first word. You are to try to anticipate the word that will be paired with the first or stimulus word. Do not be surprised if some of the words are identical to those observed during the first part of the experiment.

The Group 0 instructions were modified to accommodate the lack of Stage 1 instructions.

Verbal material.—The material consisted of a selection of the five-letter, low frequency-of-usage words employed by Horton and Kjeldergaard (1961). The lists were constructed so that the three related A, B, and C words of each association had minimal similarity. The six A-B, A-C associations were divided randomly into three sets, each of two associations. A 3 × 3 randomly permuted Latin square was employed for the assignment of the three word sets to the three Stage 2 conditions (B-C, B-Cr, B-D) so that each one third of the *S*s had one of the three sets of verbal material assigned to one of the three Stage 2 experimental conditions.

A second verbal material control was the reversal of the Stage 2 stimulus and response words. Since the Stage 1 B and C response words occurred with equal frequency, the subsequent assignment of the B and C words to stimuli or responses of Stage 2 was arbitrary. Hence, for one half of the *S*s, the stimulus and response words of Stage 2 were reversed. This control was orthogonal to the 3 × 3 Latin square material control and the three Stage 1 degree of training conditions. The *S*s were college students in introductory psychology who were assigned to the experimental conditions by use of a table of random numbers.

RESULTS

The data analyzed consisted of the sum of the correct responses for the two associations of each response condition (B-C, B-Cr, B-D).

Preliminary analyses.—An analysis of variance performed on the Stage 2 data revealed an $F < 1.00$ for the three sets of Stage 2 verbal material and for the Stage 2 stimulus-response reversal sources of variation. The data of these conditions were therefore pooled for further analyses.

An analysis of variance also revealed an $F < 1.00$ for the number of correct responses in Stage 1 (anticipations stating B or C were considered correct) for the three response conditions of Stage 2.

A significant correlation coefficient, $r\ (58) = .50$, $p < .01$, was obtained for the relation of correct responses in Stage 1 to correct responses in Stage 2 for the pooled data of Groups 40 and 80. With the lack of significance of the Stage 2 response condi-

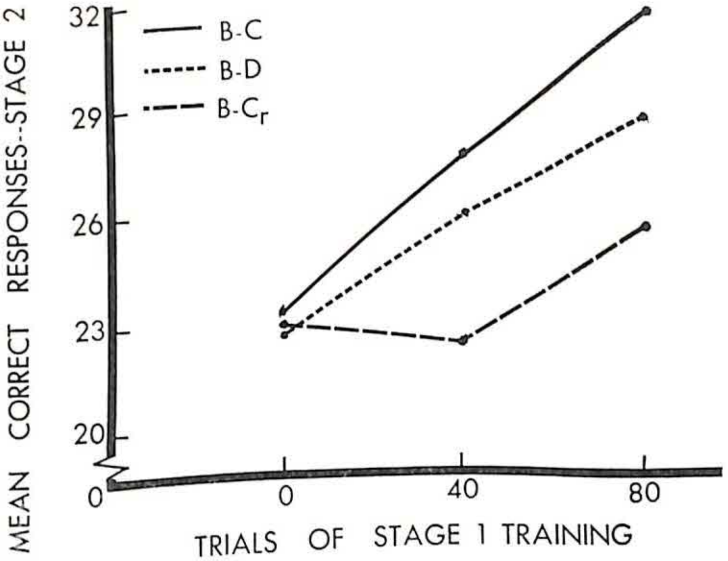


FIG. 1. Mean correct responses of Stage 2 as a function of degree of Stage 1 training for the B-C, B-Cr, and B-D conditions.

tions in Stage 1 and the finding of the significant correlation coefficient, analyses of covariance were employed for the Group 40 and Group 80 Stage 2 data which take into account correct responses in Stage 1.

Stage 2 analyses.—Figure 1 presents Stage 2 correct responses as a function of degree of Stage 1 training for the B-C, B-Cr, and B-D conditions. Although general improvement in performance occurs which is probably attributable to familiarization (Schulz & Tucker, 1962; Underwood & Schulz, 1960), differences in performance occur in the hypothesized directions for Groups 40 and 80.

An analysis of variance performed on the Group 0 data revealed that the Response Condition source of variation is not significant ($F < 1.00$). The analysis of covariance performed on the Group 40 data revealed a significant Response Condition source of variation, $F\ (2, 86) = 5.48$, $p < .01$. Subsequent analyses performed in order to detect which of the response conditions differed significantly from each other indicated a significant difference between the adjusted means of the B-C and B-Cr conditions, $F\ (1, 57) = 8.48$, $p < .01$, and a significant difference between the B-Cr and B-D adjusted means, $F\ (1, 57) = 4.86$, $p < .05$. The difference between the B-C and B-D adjusted means is not significant, $F\ (1, 57) = 1.16$.

The Group 80 covariance analysis revealed a significant Response Condition source of variation, $F\ (2, 86) = 8.95$, $p < .01$. The subsequent tests indicated significant differences between the B-C and B-Cr adjusted means, $F\ (1, 57) = 17.3$, $p < .01$, the B-C and B-D adjusted means, $F(1, 57) = 4.30$, $p < .05$, and the B-Cr and B-D adjusted means, $F\ (1, 57) = 4.30$, $p < .05$.

Transfer.—A transfer measure was employed to determine the relative amount of transfer for Groups 40 and 80. The measure was $(E_{40 \text{ or } 80} - E_0)/(C_{40 \text{ or } 80} - C_0) \times 100$, where 0, 40, and 80 represent pretraining groups, E represents either the B-C or B-Cr

TABLE 1

FREQUENCY OF INTRUSIONS FOR GROUPS 0, 40, AND 80 OF THE B-C, B-Cr, AND B-D WORDS (COLUMNS) IN THE B-C AND B-Cr CONDITIONS (ROWS)

Response Cond.	Degree of Training	B-C		B-Cr		B-D
		S	R			
B-C	0	0	28	21		31
	40	2	25	36		14
	80	1	12	23		7
				S	R	
B-Cr	0	32		1	13	42
	40	37		13	33	15
	80	33		16	37	37

Note.—See text for subcolumn explanation.

condition, and *C* represents the B-D condition. The transfer scores were as follows: 40, B-C, +126%; 80, B-C, +136%; 40, B-Cr, −12%; 80, B-Cr, +46%. Since E_0 and C_0 were virtually equal, and thus 100% implies no transfer, the results indicate that the B-C condition yields slightly increasing positive transfer and the B-Cr condition yields a decreasing amount of negative transfer as pretraining increases from 40 to 80 trials.

Intrusion data.—Table 1 presents the Stage 2 intrusion frequency data for the B-C and B-Cr response conditions of Groups 0, 40, and 80. The three main columns represent intrusions of words of the B-C, B-Cr, and B-D conditions into the B-C and B-Cr conditions (rows). The two subcolumns of the first column of the B-C row represent whether the intrusion was the stimulus (S) or response (R) of the second B-C association. The two subcolumns of the B-Cr column and row also represent whether the intrusion was the stimulus (S) or response (R) of the second B-Cr association. It should be noted that in the B-Cr condition, the two Stage 2 B-Cr associations were such that the two stimuli of Stage 2 were B-C responses of Stage 1 associated with the same stimulus (A); and, the two responses of Stage 2 were B-C responses associated with the same stimulus (A) in Stage 1.

Two aspects of the data are noted. First, for B-C, only 3 intrusions were made by the stimulus item of the second B-C association; however, for B-Cr, the stimulus word of the second B-Cr association yielded 29 intrusions in Groups 40 and 80 and only 1 in Group 0. Second, for B-Cr, there is a relatively large intrusion frequency in Groups 40 and 80 of the response word of the second B-Cr association.

DISCUSSION

Comparison of the results of the response conditions of Groups 0, 40, and 80 supports the existence of positive and negative mediational effects in simultaneous A-B, A-C acquisition. Moreover, the covariance and transfer analyses indicate that such effects occurred when correct responses on Stage 1 and the improvement in performance of B-D as a function of pretraining were considered.

The present findings may be interpreted in terms of a modification of Postman's (1961) transfer-retroaction theory. Briefly, this view is that with high B-C similarity, A-B is mediated in subsequent A-C training in this manner: A-(B)-C; on the other hand, with low B-C similarity, A-B competes with A-C and A-B is unlearned. In the present experiment, where B and C are related by common association to A during Stage 1 rather than by formal similarity, it may be theorized that B-C strength develops in Stage 1 due to the implicit statement of B during A-C acquisition: A-(B)-C, and correspondingly, A-(C)-B. The present data further suggest that Stage 1 B-C strength increases with Stage 1 training. However, when a new response is learned in Stage 2 to the B stimulus, the mediated B-C association first must be extinguished. The B-Cr intrusion analysis supports this view since the frequency of the Stage 1 C items associated with the B stimulus words of Stage 2 was considerably greater for Groups 40 and 80 than Group 0 or the B-C groups. Thus far, this explanation provides for the general superiority of B-C performance to B-D and B-Cr.

An additional factor that produces poorer acquisition in the B-Cr condition than in the B-D or B-C conditions is that the B-C items of a given Stage 1 association both appeared as stimuli or responses in Stage 2 in the B-Cr condition. The intrusion analysis indicates a relatively large frequency of the Stage 2 response that was associated with the same stimulus in Stage 1. Thus, not only must extinction occur to the B-C association of Stage 1, but the interference produced by the response item associated with the same stimulus item of Stage 1 must be reduced. This notion, however, does not provide for the relative improvement in B-Cr performance of Group 80 over Group 40.

It should be noted that according to the above considerations, the mediational effects of Stage 1 occur only between B and C. The so-called negative mediational effects are negative transfer effects based upon extinction of Stage 1 B-C associative strength during Stage 2 and the difficulty of acquisition of the new Stage 2 association.

One prediction generated by the mediation notion is that as Stage 1 A-B, A-C probability ratio is varied from .5–.5 to 1.0–0.0, positive B-C transfer effects decrease. This prediction is based upon the premise that A-B and A-C associative strength equivalence decreases over this range and A-(B)-C or A-(C)-B mediation is stronger with equivalent response strengths.

REFERENCES

Goss, A. E., & Sugarman, M. E. Paired-associates learning with varying relative percentages of occurrence of alternative response members. *J. exp. Psychol.*, 1961, **62**, 24–34.

Horton, D. L., & Kjeldergaard, P. M. An experimental analysis of associative factors in mediated generalizations. *Psychol. Monogr.*, 1961, **75** (11, Whole No. 515.)

Norcross, K. J., & Spiker, C. C. Effects of mediated associations on transfer in paired-associate learning. *J. exp. Psychol.*, 1958, **55**, 129–134.

Postman, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961. Pp. 152-179.

Schulz, R. W., & Tucker, I. F. Supplementary Report: Stimulus familiarization in paired-associate learning. *J. exp. Psychol.*, 1962, **64**, 549–550.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

Voss, J. F., Thompson, C. P., & Keegan, J. H. Acquisition of probabilistic paired associates as a function of S-R_1, S-R_2 probability. *J. exp. Psychol.*, 1959, **58**, 390–399.

(Received February 4, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 72–75

EFFECT OF CERTAIN NOISES UPON DETECTION OF VISUAL SIGNALS[1]

WILLIAM H. WATKINS[2]

Decision Sciences Laboratory, Electronic Systems Division, Bedford, Massachusetts

The influence of auditory noise stimulation upon detection of a visual signal was investigated by manipulation of 2 variables: noise type and mode of presentation. 6 trained Os each performed 1,200 forced-choice trials in which they were required to select 1 of 4 temporal observation intervals as that in which the visual signal occurred. It was found that detection performance was substantially better when noise was presented only during the observation intervals than when it was continuously present. Steady white noise and auditory flutter, both at 75 db. SPL, were equally effective in the apparent facilitation of visual detection.

This investigation tested for differential effects of four forms of acoustic noise stimulation upon visual signal detection. A forced-choice method of obtaining response data was used in preference to older psychophysical methods, many of which have been criticized with regard to assumptions concerning high thresholds and chance factor determination of responses to subthreshold stimuli (Swets, Tanner, & Birdsall, 1961).

Maier, Bevan, and Behar (1961) have cited recent studies which reported effects to which the term intersensory facilitation has been applied. Experiments of this class have typically involved observations of slightly heightened perceptual experience, often in the form of measured variables such as cff, with the improvement attributable to auxiliary acoustic stimulation. Several experiments (Chapanis, Rouse, & Schachter, 1949; Inaba, 1957; Maier et al., 1961; Walker & Sawyer, 1961) have been reported, however, in which presumably "favorable" conditions did not yield positive facilitatory findings. Broadbent (1958) and Licklider (1961) have summarized some of the abundant intersensory inhibition evidence. Also relevant are Hernandez-Peón (1961) and Harris (1950).

METHOD

Six trained *Os* were tested at a visual detection task while receiving controlled auditory input. Testing was accomplished in sets of 240 trials, equally distributed among four acoustic conditions. Thirty trials under a particular acoustic condition constituted a *run*, lasting 3.75 min. A *set* consisted of two runs per condition for each of the four conditions, or eight runs. These eight runs were randomly ordered within each set, which usually consumed a 1-hr. testing session. Rest periods were at least 2 min. between runs and generally 24 hr. between sessions. The *O* was informed of his score for a run (number correct of 30 trials) at the completion of that run.

Noises.—The output of a white noise generator, passed through a 100–3,000 cycle filter, and recorded on magnetic tape, was played through a reproducing circuit terminating in binaural Permoflux PDR-20 earphones. The overall sound pressure level (SPL) was 75 db. at each earphone. The

---

[1] This research was accomplished at the University of Texas as part of a PhD thesis in psychology. The author is indebted to Robert K. Young, who served as Dissertation Committee Chairman. This report is also identified as Electronic Systems Division Technical Documentary Report No. ESD-TDR-63-202. Furthur reproduction is authorized to satisfy needs of the United States Government.

[2] Major, United States Air Force.

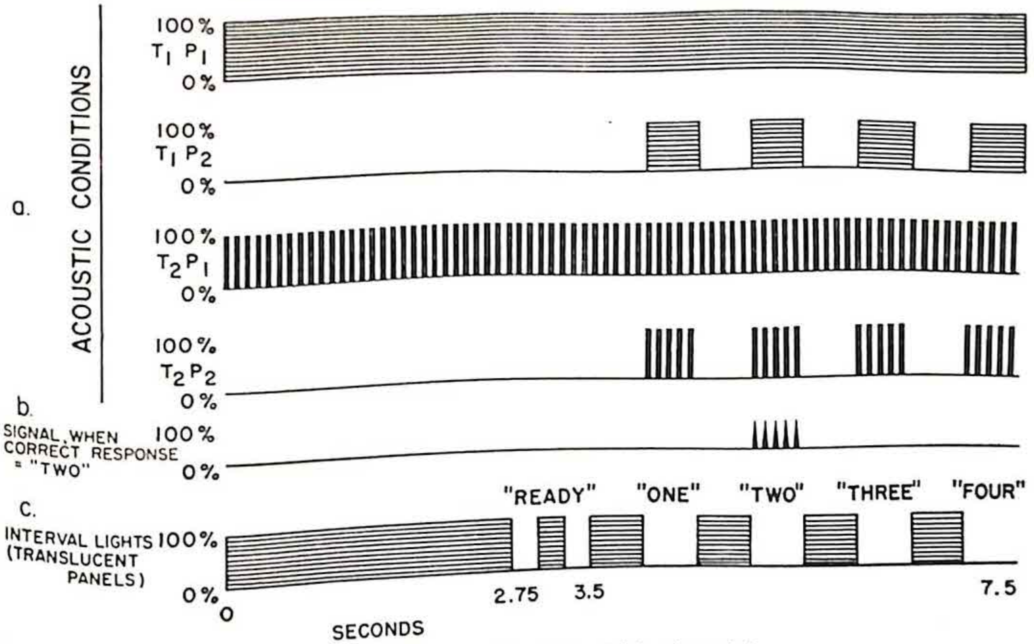


FIG. 1. Sequence of events within the trial.

noise was manipulated in terms of its temporal characteristics. Type 2 (T_2) noise consisted of 10 burst per second auditory flutter, with each burst and each separation lasting 50 msec. Type 1 (T_1) noise was not interrupted except as mentioned below; it was "smooth" in comparison to Type 2. At a different level of consideration, interruption of noises was employed to allow *presentation* either continuously during the run (P_1) or only during the four observation intervals of each trial (P_2). Figure 1 depicts the acoustic conditions.

Visual environment.—From a distance of 1 m., *O* viewed binocularly a flat-textured circular patch of illuminated white glass, 17 cm. in diameter, having the intense luminance of 625 ft-L. An opaque tube fitted with a cushioned face piece limited peripheral photic stimuli to the black interior of the tube. Visual signals were presented at an otherwise darkened spot in the center of the white circle. Signals were produced by a Model PS 2A photo stimulator (Grass) clamped to the back of the viewing apparatus. The size of the aperture which permitted passage of the signal light was adjustable around a median diameter of 1.85 mm. The aperture for each *O* was so selected as to achieve roughly equivalent detection among *O*s. One change of aperture "within an *O*" was made; it occurred after the second complete set (sixteenth run) and was to the next smaller setting. Signals always consisted of a $\frac{1}{2}$-sec. series of flashes. The separate flashes occurred with a frequency of 10/sec and each flash had a duration of approximately 10 μsec. The standard sequence of events within a trial was defined by illumination and darkening of translucent panels, 6 cm. wide, located just above and below the target patch.

Timing of events.—The temporal pattern of each trial was controlled by microswitches, operated by revolution of a motor-driven cam wheel assembly. Figure 1 shows the time schedule employed. The *O* was always presented with one of the four acoustic conditions, as diagramed in the top four tracks, a visual signal to be detected, as exemplified in the fifth track, and the interval light sequence shown. The $2\frac{3}{4}$-sec. interval which followed Observation Interval 4 served for response recording and intertrial rest. Under the T_2 noise conditions, the photo stimulator was externally triggered by pulses originating in the noise gate device. This enabled presentation of the photic signals in flutter-flash phase with the separate bursts of noise.

Subjects and procedure.—The six *O*s averaged 23 yr. of age. Four wore eyeglasses, and one (ES) claimed severe hearing loss monaurally. All were individually trained to the point of approximately level performance at some aperture. This training experience was in all cases distributed equally among the four acoustic conditions, and typically con-

TABLE 1

Analysis of Variance Summary

Source	*df*	*MS*	*F*
Between *O*s (O)	5		
Noise presentation mode (P)	1	650.096	7.58*
Noise type (T)	1	1.504	.17
Run sequence (R)	9	40.871	1.65
O × P	5	85.696	
O × T	5	8.974	
O × R	45	24.796	
P × T	1	.504	.35
P × R	9	7.661	.91
T × R	9	6.726	.68
O × P × T	5	1.434	
O × P × R	45	8.405	
O × T × R	45	9.885	
P × T × R	9	7.078	.99
O × P × T × R	45	7.164	
Total	239		

* $p < .05$.

sumed nine 1-hr. sessions. Following the training, 40 experimental runs were accomplished by each *O*. Responses were registered on moving paper tape according to the one among four keys *O* depressed after Observation Interval 4.

Results and Discussion

The results of an analysis of variance of the 240 experimental runs are presented in Table 1. For the group, visual signal detection was superior when noise was presented simultaneously with the observation intervals to performance under continuous noise presentation conditions. No other variable was found significant, nor were there any significant interactions. The O × P *MS* was used to test the P effect, in line with the assumption that O represented a random variable. It was not conclusively established that this assumption of random *S* selection was justified. If O were in fact a fixed variable, the proper error term in the entirely fixed model would be O × P × T × R. Mean scores are presented by *O* and by condition in Table 2. Chance level detection in four-alternative-choice type trials is, of course, a score of 7.5 for 30 trials. The superiority of the P_2 over the P_1 mean amounted to 3.30 more correct identifications of signal intervals per run—i.e., approximately 20%.

Group means are plotted for the four acoustic conditions along the five successive sets in Fig. 2. Here, the comparatively wide separation of points associated with P_1 (dots) from those of P_2 (circles) is evident in relation to the insignificant T_1 (solid lines)-T_2 (dashed lines) difference.

This experiment has affirmed that acoustic stimuli can function to improve visual signal detection. It is apparent, also, that the temporal characteristics of the sound may be of critical importance to the facilitatory potential of white noise. No conclusions are possible regarding *how* the difference in noise pres-

TABLE 2

Mean Number Correct Trials (in Runs of 30 Trials)

O	T_1P_1	T_2P_1	T_1P_2	T_2P_2	P_1	P_2	T_1	T_2	All Conditions
BY	17.1	17.1	15.9	16.5	17.1	16.2	16.5	16.8	16.65
HG	11.9	9.8	17.0	16.1	10.8	16.5	14.4	12.9	13.70
ES	13.3	14.6	20.8	21.4	13.9	21.1	17.0	18.0	17.52
MB	20.1	20.2	21.6	22.4	20.2	22.0	20.8	21.3	21.07
JS	16.0	17.3	20.3	21.1	16.6	20.7	18.2	19.2	18.68
JG	19.7	19.5	21.7	21.3	19.6	21.5	20.7	20.4	20.55
Group averages	16.35	16.42	19.55	19.80	16.38	19.68	17.95	18.11	18.03

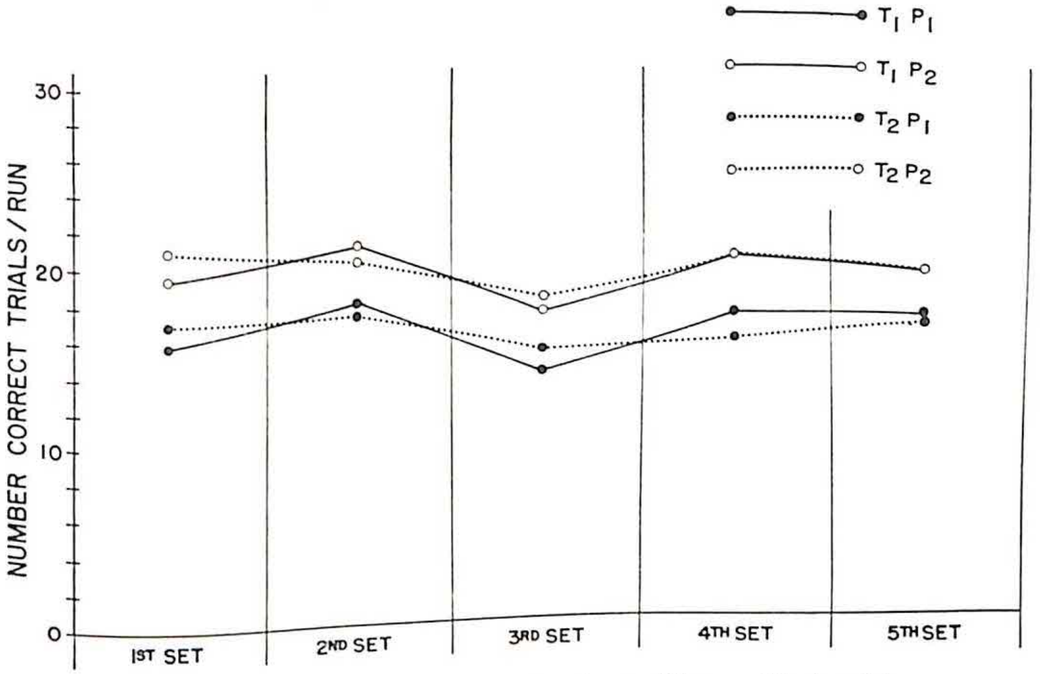


FIG. 2. Mean scores by sets and acoustic conditions. (Each point represents two runs by each of six *Os*.)

entation modes operated to cause the effect demonstrated.

Ogilvie (1956) found visual performance as expressed by cff to be facilitated more by in-phase auditory flutter than other forms of white noise. In the present study, visual signal detection was independent of the type noise presented: steady or flutter, in-phase with light flashes. There is limited correspondence between the findings of these two investigations: that the outcomes are at all similar may be noteworthy in view of the sharply divergent techniques employed.

When the effect described in this study has been accounted for neurophysiologically, it may be found that processes and mechanisms are involved which also serve to mediate in various phenomena of "attention" and "time perception."

REFERENCES

BROADBENT, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

CHAPANIS, A., ROUSE, R. O., & SCHACHTER, S. The effect of intersensory stimulation on dark adaptation and night vision. *J. exp. Psychol.*, 1949, **39**, 425–437.

HARRIS, J. D. *Some relations between vision and audition.* Springfield, Ill.: Charles C Thomas, 1950.

HERNANDEZ-PEÓN, R. Reticular mechanisms of sensory control. In W. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961. Pp. 497–520.

INABA, K. Investigations of sensory interaction. Unpublished doctoral dissertation, Purdue University, 1957.

LICKLIDER, J. C. R. On psychophysiological models. In W. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961. Pp. 49–72.

MAIER, B., BEVAN, W., & BEHAR, I. The effect of auditory stimulation upon the critical flicker frequency for different regions of the visible spectrum. *Amer. J. Psychol.*, 1961, **74**, 67–73.

OGILVIE, J. C. Effect of auditory flutter on the visual critical flicker frequency. *Canad. J. Psychol.*, 1956, **10**, 61–68.

SWETS, J. A., TANNER, W. P., & BIRDSALL, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 301–340.

WALKER, E. L., & SAWYER, T. M. The interaction between critical flicker frequency and acoustic stimulation. *Psychol. Rec.*, 1961, **11**, 187–193.

(Received February 9, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 76–79

CLUSTERING AS A FUNCTION OF RESPONSE DOMINANCE[1]

W. A. BOUSFIELD AND C. R. PUFF

University of Connecticut

This experiment was designed to test the predicted effectiveness of response dominance as assessed by Underwood and Richardson for inducing clustering in free recall. A randomized list of 24 words including 8 items representing a given concept at a high dominance level was presented to 34 Ss. For 24 Ss, 8 words representing the given concept at a low dominance level replaced the high dominance words of the 1st list. Significant clustering was found only for the high dominance words. A new formula was used for assessing the incidence of clustering.

Underwood and Richardson (1956a) introduced the term response dominance to denote the relative strength of the tendency of *S*s to give a specified type of response. Working with restricted associative responses, their method was to present nouns as stimuli with instructions to *S*s to respond only with sense impressions. The relative dominance of the associative responses, which were in the form of adjectives, was indicated by the cultural frequencies of their occurrence. Since these adjectival responses were conceptual in nature, Underwood and Richardson (1956b) used the normative data they had obtained earlier for a study of concept learning. They showed that the rate of learning of specified sense impressions as responses to given nouns was a positive function of the dominance levels of the responses. From the standpoint of the present study the important feature of the Underwood and Richardson (1956a) normative responses is that due to the restrictions under which they were obtained, they are in effect relatively pure stimulus equivalences. In view of the tenability of the proposition that stimulus equivalence may be regarded as a defining characteristic of conceptually related words, Underwood and Richardson's norms were especially well suited for the study of concept formation. Bousfield (1953) showed that groups of stimulus words which are related on the basis of taxonomic concepts—e.g., COW, BEAVER, BEAR, LION—when presented to *S*s in a randomized list, tended to cluster in free recall. That is, *S*s tended to recall the items in runs of two or more in the same taxonomic class. Free associational norms show that words in the same taxonomic class tend to elicit common responses. In the foregoing example, all the stimulus words elicit ANIMAL as a common response. These stimulus words are therefore equivalent stimuli. In addition, groups of taxonomically related words of the type commonly used for inducing clustering typically have direct associative connections with each other. Bousfield, Steward, and Cowan (1961) found an appreciable positive relationship between interitem associative strength as measured by Deese's (1959) method, and the clustering obtained from using taxonomic categories and words given

---

[1] This is Technical Report No. 39 under Contract Nonr-631(00) between the Office of Naval Research and the University of Connecticut. Reproduction in whole or in part is permitted for any purpose of the United States Government.

TABLE 1

SUBLISTS OF STIMULUS WORDS WITH CONCEPTS AND DOMINANCE LEVELS (DL)

Sublist 1 (High DL; Concept, Smelly)		Sublist 2 (Low DL; Concept, Smelly)		Sublist 3 (Medium DL; Concept, Round)		Sublist 4 (Medium DL; Concept, Soft)	
Words	DL	Words	DL	Words	DL	Words	DL
AMMONIA	88	SARDINE	30	BALLOON	55	VELVET	67
MANURE	83	HOSPITAL	23	BELLY	43	FLANNEL	54
GARBAGE	80	HOG	15	PLATTER	38	MOCCASIN	53
SKUNK	78	VINEGAR	14	DERBY	33	SKIN	42
GARDENIA	65	DAFFODIL	12	POT	29	KITTEN	41
SEWER	61	COFFEE	12	BRACELET	25	JELLYFISH	31
GARLIC	58	DIAPER	7	APPLE	19	CUSTARD	26
CHEESE	25	CINNAMON	6	HELMET	17	LIPS	24
Mean DL	67.25	Mean DL	14.88	Mean DL	32.38	Mean DL	42.25

Note.—List A consists of Sublists 1, 3, and 4; List B consists of Sublists 2, 3, and 4.

in the cultural norms of Cohen, Bousfield, and Whitmarsh (1957). In so far as such direct associational relationships exist they necessarily contribute to the incidence of clustering. In the present study of clustering such interitem associations were minimized by using the Underwood and Richardson norms for selecting conceptually related words. This provided for a test of the application of the stimulus equivalence paradigm in a relatively pure form. The following two hypotheses were tested: (*a*) Response dominance can serve as a basis for inducing clustering in free recall. (*b*) The greater the dominance level of a common conceptual response to a group of words, the greater the amount of clustering of these words in free recall.

METHOD

Subjects.—The *S*s were two classes of undergraduate students in psychology comprising 34 and 24 members, respectively.

Stimulus materials.—The stimulus words comprised two 24-word lists of nouns selected from the response dominance norms of Underwood and Richardson (1956a). The lists, designated, respectively, as A and B, in turn comprised three of the eight-word sublists shown in Table 1. List A contained Sublist 1 with high DL (dominance level) words for the concept *smelly*; Sublist 3 with medium DL words for the concept *round*; Sublist 4 with medium DL words for the concept *soft*. List B contained Sublist 2 with low DL words for the concept *smelly*; Sublists 3 and 4 which were common to both List A and List B. Each word representing the concept of a given sublist had no ratings representing the remaining concepts. Since the plan of the experiment involved three separate presentations of each list for learning, three separate randomized serial orders of the lists were prepared.

Procedure.—List A was presented to the group of 34 *S*s and List B to the group of 24 *S*s. The *E* first distributed 8½ × 11 in. blank sheets to *S*s for recording their recalls. He then gave the following instructions:

> I shall read a list of words to you several times. The words will be spelled as they are read the first time. When the reading of the words has been completed, you are to recall as many of the words as you can. The order in which you recall the words is of no consequence. Simply write them in the order in which they occur to you. Write the words in a single column down the left side of the page.

During the first reading, *E* pronounced and spelled the words at a rate of about 4 sec. per word. For the second and third presentations, *E* read the words at a rate of about 2 sec. per word. A period of 5 min. was allowed for recall.

RESULTS

In analyzing the data the question arose regarding the most appropriate method for evaluating the incidence of clustering within individual concept categories. The conventional unit of measurement of clustering is the repetition which is defined as the sequential occurrence of two items in the same category. In these terms, the number of repetitions in a run or cluster of items in the same category is the number of items minus one. In the present study, the model chosen for the derivation of a formula for the number of repetitions to be expected on the basis of chance involved two major assumptions: first, when *S* undertakes recall, a certain number of items are not available and therefore have no effect on the incidence of clustering; second, at any stage of recall all the words remaining to be recalled are equally available, and are chosen without replacement.[2] For the analysis of the data of the present study the model was applied so as to give the number of repetitions expected on the basis of chance in a single category of the recalled items. Suppose *S* recalls a total of n items from a list of words used for inducing clustering. Of these items a total of m_k is from Category k. The question is that of how many repetitions should occur by chance in Category k under these conditions. For a given n, the larger the m_k the more repetitions in Category k we should expect on the basis of chance. The clustering regarded as critical is that which exceeds this chance amount. From the standpoint of theory, clustering measured on this basis would appear to be attributable to what Underwood and Schulz (1960, p. 92) have termed association or hook-up rather than to response recall. Though these assumptions are obviously open to criticism, they provide a serviceable approximation for a null hypothesis. According to the indicated model the formula for $E(r_k)$, the expected repetitions in a given category k is:

$$E(r_k) = \frac{m_k(m_k - 1)}{n}$$

Thus, if the number of items recalled in Category k is 6, and if a total of 15 items is recalled, the number of expected repetitions in Category k is 2.

The first steps undertaken in the analysis of the data were to make three counts from the data of each *S*: (*a*) total number of items recalled; (*b*) number of critical concept, smelly, words recalled; (*c*) number of repetitions of critical concept words. Applying the formula for $E(r_k)$, the number of expected repetitions in the critical category was then computed for each *S*. With these data, paired frequency distributions were compiled for the 34 *S*s who received List A with the high DL smelly words, and for the 24 *S*s who received List B with the low DL smelly words. One distribution was that of the obtained repetitions, and the other comprised the expected repetitions. Three t tests of differences between paired means of the distributions yielded the following results. The difference between the obtained and expected means for the high DL words was significant, $t(33) = 3.17$, $p < .01$. The corresponding difference for the low DL words was not significant, $t(23) = .223$, $p > .05$. The difference between the number of repetitions above chance for the high DL words and for the low DL words fell slightly short of the .05 level, $t\ (56) = 1.98$, $p > .05$.

---

[2] This model was developed for the present study by A. K. Bousfield. It differs from that of Cohen, Sakoda, and Bousfield (1954) who used the assumption of drawing without replacement from a pool comprising all the items of the given list.

DISCUSSION

The results provide confirmation for the first experimental hypothesis which states that response dominance as appraised by Underwood and Richardson (1956a) can serve as a basis for inducing clustering in free recall. The results failed, however, to confirm the second hypothesis which states that response dominance should serve as a parameter of clustering. The results appear to show that stimulus equivalence by itself can be made to serve as a basis for inducing clustering. The importance of stimulus equivalence for clustering was demonstrated in a different way by Bousfield, Steward, and Cowan (1961). They undertook the testing of two predictive measures of clustering which were based on the use of free associational norms. As already noted, Deese's (1959) measure of interitem associative strength was found to have a positive relationship with the clustering obtained through the use of taxonomic categories of items. On the other hand, a measure of stimulus equivalence appeared to correlate more highly than interitem associative strength with the incidence of clustering. It is evident that both interitem association and stimulus equivalence contribute to the clustering. Marshall and Cofer (1963) in their review of measures of word relatedness distinguished between what they termed direct and nondirect association. In effect these are alternative terms for recognizing the distinction emphasized in the present study. As they have implied, the important problem is not that of which of the two types of association is more important, but that of assessing their relative effects in various learning processes.

REFERENCES

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, **49**, 229–240.

BOUSFIELD, W. A., STEWARD, J. R., & COWAN, T. M. The use of free associational norms for the prediction of clustering. Technical Report No. 36, 1961, University of Connecticut, Contract Nonr-631(00), Office of Naval Research.

COHEN, B. H., BOUSFIELD, W. A., & WHITMARSH, G. A. Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957, University of Connecticut, Contract Nonr-631(00), Office of Naval Research.

COHEN, B. H., SAKODA, J. M., & BOUSFIELD, W. A. The statistical analysis of the incidence of clustering in the recall of randomly arranged associates. Technical Report No. 10, 1954, University of Connecticut, Contract Nonr-631(00), Office of Naval Research.

DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312.

MARSHALL, G. R., & COFER, C. N. Associative indices as measures of word relatedness: A summary and comparison of ten methods. *J. verbal Learn. verbal Behav.*, 1963, **1**, 408–421.

UNDERWOOD, B. J., & RICHARDSON, J. Some verbal materials for the study of concept formation. *Psychol. Bull.*, 1956, **53**, 84–95. (a)

UNDERWOOD, B. J., & RICHARDSON, J. Verbal concept learning as a function of instructions and dominance level. *J. exp. Psychol.*, 1956, **51**, 229–238. (b)

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received February 11, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 80–82

REVERSAL OF CONDITIONED DISCRIMINATION OF THE EYELID RESPONSE[1]

C. MICHAEL LEVY, DAVID A. GRANT, AND ALTON H. CLARK

University of Wisconsin

40 Ss were given 60 differential conditioning trials, ½ with the positive stimulus, ½ with the negative, using the eyelid CR to light based on the reflex wink to an air puff. Following differential training 30 additional trials were given, and for ½ of the Ss the positive and negative stimuli were reversed. Good discrimination and discrimination reversal were obtained, although not as rapid as in an analogous verbal conditioning situation. Reversal from negative to positive CS was not as complete as reversal from positive to negative.

The problem was to investigate the reversal of conditioned discrimination in the eyelid CR. Although discrimination reversal has long been an interesting problem with instrumental responses (e.g., Harlow, 1949), the phenomenon has rarely been investigated in the area of classical conditioning. Scattered reports of the reversal of conditioned discrimination are available from the Russian literature, but they are usually reports of experiments carried out in connection with the investigation of other problems, such as the activity of the nervous system and neurasthenia (e.g., Belyayeva, 1961). The finding of Hartman and Grant (1962) that excellent differential conditioning of the eyelid response could be obtained when a CS-UCS interval of 800 msec. was used made it feasible to investigate the reversal of discrimination with this response.

METHOD

Apparatus.—The standard Wisconsin eyelid conditioning apparatus as modified for conditioned discrimination by Hartman and Grant (1962) was used. The *S*s faced a rectangular enclosure that was painted flat white. Approximately 80 cm. from *S* were two milk-glass disks, 10 cm. in diameter set 15 cm. apart and slightly below his normal line of sight. Ambient illumination was approximately 1.00 mL. and the CS consisted of a 0.8-mL. increase in brightness of one of the milk-glass windows. The UCS, a corneal air puff, lasted 50 msec. and its intensity was regulated by a 150-mm. column of mercury. The CS lasted 1 sec. and its onset was followed 800 msec. later by the onset of the UCS. Intertrial intervals varied from 10 to 40 sec. as regulated by a Western Union tape transmitter.

Procedure.—All *S*s were given 60 discrimination training trials, 30 reinforced trials with one stimulus light reinforced and 30 trials with the other stimulus light which was never reinforced. For half the *S*s the left light was reinforced, and for the other half of the *S*s the right light was reinforced.

After the 60 discrimination trials, half of the *S*s, the Control group, continued as before for 30 more trials, and the other half of the *S*s, the Experimental group, had the conditioned stimuli reversed so that the previously reinforced CS was no longer reinforced and the previously unreinforced CS was now regularly reinforced. The Reversal group, thus, received discrimination reversal trials following discrimination training. The positive and negative trials were randomly interspersed with the restriction that Trials 1, 60, 61, and 90 were reinforced, and half of the first 60 and half of the last 30 trials were reinforced.

Subjects.—Twenty men and 20 women volunteers from the introductory psychology

---

[1] This research was supported in part by the Research Committee of the Graduate School of the University of Wisconsin with funds provided by the Wisconsin Alumni Research Foundation and in part by Research Grant MH 06792 from the National Institutes of Health.

course at the University of Wisconsin served as *S*s. The sex ratio was balanced in the control and reversal groups, but *S*s were otherwise unsystematically assigned. The *S*s were not familiar with eyelid conditioning and were read the neutral instructions used in the Wisconsin laboratories. None of the *S*s was informed that the configuration of stimulus lights might be reversed.

RESULTS

Eyelid responses greater than 1 mm. as recorded, which occurred with latencies between 250 and 800 msec. after the onset of the CS, were scored as CRs. Only five *S*s, two in the Reversal and three in the Control group, were classified as Voluntary Responders by the time derivative criterion (Hartman & Ross, 1961) and their data were retained with the rest of the data.

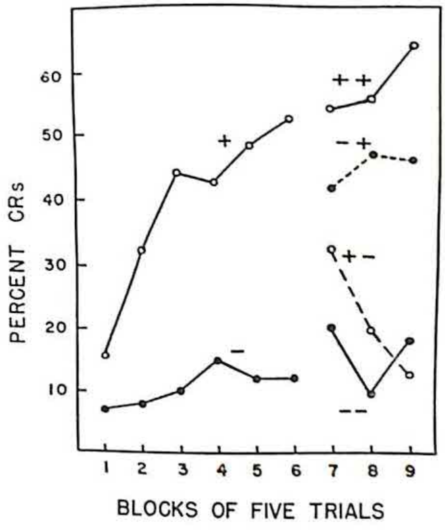


FIG. 1. Percent CRs to reinforced (+ and unreinforced (−) stimuli during sucessi five-trial blocks of the experiment. (Tl −+ and +− curves are for stimuli for whic reinforcement conditions had been reverse during the discrimination reversal phase the experiment, the latter symbol referrin to the actual reinforcement condition durin reversal.)

The principal results of the experiment are shown in Fig. 1 where percent CRs is plotted for successive 5-trial blocks. The first 6 of the 5-trial blocks were discrimination training and Blocks 7, 8, and 9 involve the control and discrimination reversal trials. It is obvious from the figure that a high degree of differential conditioning had occurred after 30 reinforced and 30 unreinforced trials, $F\,(1, 38) = 39.59$, Error $MS = 13.01$. There was no difference between the control and reversal groups in the discrimination training ($F < 1.0$). On the reversal blocks the control procedure produced a continuation of the positive and negative curves, the ++ and −− curves of Fig. 1. The curves of the Reversal group were designated −+ and +− where the first sign refers to the positive or negative character of the CS during differential conditioning, and the second sign refers to the character of the CS during reversal. The discrimination reversal occurred fairly rapidly, but was more complete in the reversal from positive to negative than fror negative to positive. The differenc between the Reversal and Contro groups was highly significant, $F\,(1, 38$ $= 7.79$, Error $MS = 10.58$. Actually a trial by trial plot of the first 5 trial of the Discrimination Reversal Training Block 7 showed that considerabl reversal had occurred, on the average by the third discrimination reversa trial.

DISCUSSION

Although good conditioned discrimination and discrimination reversal wer obtained, using the eyelid response neither the discrimination nor the reversal was as rapid or as complete as car be obtained with a verbal instrumenta response. For example, data were obtained from 55 men and women *S*s in a noncontingent "guessing game" (Humphreys, 1939) where the situation paralleled as closely as possible the relation ships in the present experiment. Two "CS" lights were placed at the top of a Masonite board on the left and right, and in the center, below, was a "UCS" light covered by a green filter. In discrimination training one of the "CS" lights was always followed by illumination of the "UCS" light but the other was never "reinforced." After 60 training trials the positive and negative "conditioned stimuli" were reversed. By the twelfth training trial *S*s were responding positively 90% or more to the positive "CS" and 10% or less to the negative "CS." By the fourth reversal trial *S*s were responding more than 95% of the time to the new positive stimulus and less than 5% of the time to the new negative stimulus. The discrimination and its reversal clearly occur more rapidly and more completely with the verbal instrumental response than with the conditioned eyelid response.

The asymmetry in discrimination reversal, indicated by the fact that the frequency of responses to the −+ stimulus is clearly less than the frequency of response to the ++ stimulus whereas the level of response to the +− stimulus drops to the level of the −− stimulus, is of considerable interest. Although this asymmetry may be due to the relative slowness of acquisition as compared with extinction of the eyelid CR, it certainly parallels the persistence of inhibition from formerly negative stimuli found in instrumental learning situations by other investigators (Hayes, 1953; Ross, 1962a, 1962b, 1962c). Pavlov (1927, pp. 196–200) has also commented on the persistence of inhibition in discrimination reversal, and the phenomenon is also occasionally reported by other investigators (e.g., Yoshii, Matsumoto, Ogura, Shimokochi, Yamaguchi, & Yamasaki, 1960, p. 201). In his discussion of the phenomenon, Pavlov attributed it to the presence of interspersed trials with the new negative CS, but this interpretation has not been intensively investigated, so that it remains one of the interesting asymmetries between excitatory and inhibitory processes.

REFERENCES

Belyayeva, Z. V. Characteristics of the basic nervous process in neurasthenia. *Zh. nervn. Deiatel.*, 1961, **11**, 30–36.

Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

Hartman, T. F., & Grant, D. A. Differential eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1962, **64**, 131–136.

Hartman, T. F., & Ross, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *J. exp. Psychol.*, 1961, **61**, 334–338.

Hayes, K. J. The backward curve: A method for the study of learning. *Psychol. Bull.*, 1953, **60**, 269–275.

Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep) London: Oxford Univer. Press, 1927.

Ross, L. E. The effect of equal reinforcement of the positive and negative discriminanda of a learned discrimination. *J. comp. physiol. Psychol.*, 1962, **55**, 260–266. (a)

Ross, L. E. Note: Persisting inhibitory effect found in both classical and instrumental learning situations. *Psychol. Rep.*, 1962, **11**, 691–692. (b)

Ross, L. E. The response to previous discriminanda during the learning of a new problem. *J. comp. physiol. Psychol.*, 1962, **55**, 944–946. (c)

Yoshii, N., Matsumoto, J., Ogura, H., Shimokochi, M., Yamaguchi, Y., & Yamasaki, H. Conditioned reflex and electroencephalography. In H. H. Jasper & G. D. Smirnov (Eds.), Moscow colloquium on electroencephalography of higher nervous activity. *EEG clin. Neurophysiol. Supple., Amsterdam*, 1960, No. 13.

(Received February 15, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 83–90

KINESTHETIC FIGURAL AFTEREFFECTS: SATIATION OR CONTRAST

JOSEPH J. MOYLAN[1]

University of Dayton

3 experiments are reported. Experiment I investigated kinesthetic figural aftereffects (KFAE) as a function of hand satiated. Results indicated that KFAE could be elicited in the direction traditionally described, or in an opposite direction depending upon whether the satiated hand, after satiation, proceeds to the T or V bar. Experiment II replicated Experiment I with some alterations in manipulanda. Experiment III questioned the traditional satiation procedures, as well as the assumption of "satiated boundaries," prevalent in satiation theory. The data suggested that both the procedure and the concept are unnecessary.

Since the publication of Köhler and Dinnerstein's (1947) paper on kinesthetic figural aftereffects (KFAE) and the earlier Köhler and Wallach (1944) disagreement with Gibson's (1933) interpretation of aftereffect phenomenon, a number of studies investigating aftereffects have appeared (cf. Spitz, 1958). Regarding KFAE, the summary findings reported by Köhler and Dinnerstein (1947) are that the size of the T bar which is narrower than the I bar will be underestimated, a T bar wider than the I bar will be overestimated, and aftereffect will not occur when T and I bars are of equal width.

By and large, subsequent investigations of KFAE have employed essentially the same methodology as that used by Köhler and Dinnerstein (e.g., Krauskopf & Engen, 1960; McEwen & Rodger, 1960). Thus, the fingers of the satiated hand, i.e., the hand called upon to rub lightly the edges of the inspection bar or strip (I bar), were subsequently placed on a test bar or strip (T bar), the width of which *S* was to estimate, compare, or find on a variable bar (V bar) on which his unsatiated hand had been placed.

The present studies were born out of repeated experimental psychology laboratory demonstrations of KFAE. During these demonstrations students employed a wide variety of stimuli and procedures. Certain of these procedural variations appeared informally to yield results which were markedly at variance with those reported by Köhler and Dinnerstein. When students placed the satiated hand on the V bar, with the unsatiated hand on the T bar, they still demonstrated "aftereffects" but in a direction opposite to traditional aftereffect phenomena. That is, a T bar narrower than its associated I bar was judged "wider" rather than "narrower." The first study then evaluated KFAE as a function of hand satiated, to determine if differential effects obtained depending upon whether the satiated hand proceeded to the T or V bar.

EXPERIMENT I

Method

Subjects.—The *S*s were 10 right-handed male volunteers from an introductory psychology class at the University of Dayton.

---

[1] Now at Parsons College, Fairfield, Iowa.

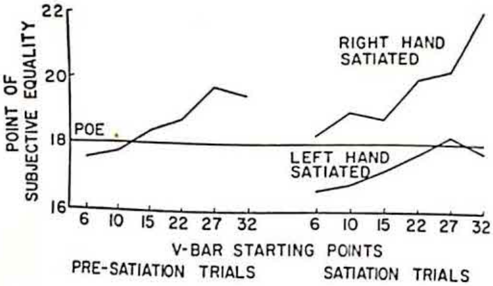


FIG. 1. Mean T bar width judgments before and after satiation (2.5-in. I bar; 1.5-in. T bar).

Apparatus.—The apparatus had three ½-in. thick pine boards: an I bar, 2.5 in. wide; a T bar, 1.5 in. wide; and a V bar, 29.5 in. long tapered from a point to 3.5 in. wide and marked off with lines ¾ in. apart. The lines were numbered consecutively beginning with No. 1 which designated the line ¾ in. from the point of the V bar. Line No. 18 was the point of equality (POE) for the 1.5-in. T bar.

Procedure.—The *S*s were brought individually into the experimental room where they were asked to rinse their hands (soap not used) in warm water and to dry them using a rough hand towel. Each *S* then made six judgments of equality with his left hand on the 1.5-in. T bar and his right hand on the V bar. On these as well as on all subsequent trials *S*'s right hand was placed on the V bar at approximately the following points: 27, 15, 6, 22, 10, and 32 providing far, medium, and near starting points above and below the POE.

After these presatiation trials *S*s were alternately assigned to right- and left-hand satiation groups. Satiation amounted to 60 sec. of rubbing the left (right) hand fingers along the top edges of the 2.5-in. bar after which the T bar and V bar were provided for a judgment. The next judgments were preceded by 20-sec. satiation periods on the I bar.

Results

The first question to be answered was whether the left- and right-hand satiation groups were equal in ability to judge the width of the 1.5-in. T bar prior to satiation trials. An analysis of variance of the T bar judgments on presatiation trials revealed that the effects of groups were not significant. Although the effects of trials were significant, $F(5, 40) = 4.44$, $p < .01$, the Groups × Trials interaction was not. Thus, of particular interest is the fact that the statistical analysis suggests that the two groups can be considered to be equal prior to satiation trials.

Since the two groups functioned identically on presatiation trials their curves have been combined for these trials in Fig. 1. Figure 1 then, clearly reveals the variation in accuracy of judgments over trials or, more specifically, according to the starting point of each trial (i.e., from the point of placement of the right hand on the V bar). Generally, hand placement below POE gave rise to a low PSE and placement above POE gave rise to a high PSE. In all probability this phenomenon is associated with the psychophysical method employed (cf. Dember, 1960, p. 39).

While actual judgment trials were given in the order described above—right hand placed at Starting Point 27 for the first judgment, 15 for the second judgment, etc.—for a clear presentation of Fig. 1, starting points are presented on the abscissa in numerical sequence. The fact that attention has not been drawn to the

TABLE 1

ANALYSIS OF VARIANCE OF RIGHT-HAND AND LEFT-HAND SATIATED GROUPS' T BAR JUDGMENTS OVER SIX SATIATION TRIALS

Source	*df*	*F*
Between *S*s		
Groups (B)	1	17.54**
Error (b)	8	
Within *S*s		
Trials (A)	5	6.15**
A × B	5	3.31*
Error (w)	40	

* $p < .05$.
** $p < .01$.

role of starting points in KFAE investigations also dictated the present form of presentation.

Table 1 summarizes the results of a second analysis of variance of data obtained on satiation trials. Results of this analysis show that *F*s for groups, trials, and interaction effects are all significant. Figure 1 graphically illustrates the differences tested. Clearly the right- and left-hand satiated groups performed differently. The significant difference over trials appears to be directly related to the starting points. Also the significant *F* for interaction seems to be due to the fact that the right-hand satiated group was more severely affected by starting points than was the left-hand satiated group. It can be seen then, that for all judgments (relative to presatiation judgments), right-hand satiated *S*s judged the T bar to be wider, while left-hand satiated *S*s judged the same T bar to be narrower.

The data of Exp. I are not simply explainable in either contrast or satiation terms. What this experiment does reveal is that T bar judgments do indeed differ depending upon whether the satiated hand proceeds to the T or V bar. Since Exp. II and III are basically methodologically refined replications of Exp. I, the results of these studies will be presented first, followed by a more detailed general discussion of all three studies.

Experiment II

Experiment I found right- and left-hand satiated *S*s giving wider and narrower T bar PSEs, respectively, than they did prior to their satiation experience. If this phenomenon is at all reliable, a simple reversal of I and T bar sizes should find left-hand satiated *S*s giving wider T bar judgments and right-hand satiated *S*s giving narrower judgments.

TABLE 2

Analysis of Variance of Right-Hand and Left-Hand Satiated Groups' T Bar Judgments over Five Satiation Trials

Source	*df*	*F*
Between *S*s		
Groups (B)	1	11.825**
Error (b)	18	
Within *S*s		
Trials (A)	4	36.674**
A × B	4	1.722
Error (w)	72	

** $p < .01$.

Method

The *S*s were 20 students drawn from the same population as in Exp. I. Changes in procedure were as follows: First, the wedge-shaped V bar method of measurement involves several serious shortcomings. Precision is lacking in that *S*'s thumb and forefinger are not necessarily parallel and *E* is frequently left with a decision as to which $\frac{3}{4}$-in. space to assign as *S*'s position. Not only is precision lacking but the very fact that *S* must feel the edges of the V bar as he seeks PSE, potentially involves some kind of satiation effect (Köhler & Dinnerstein, 1947; Krauskopf & Engen, 1960). In order to overcome these deficiencies, we turned to metric measurements and a vise arrangement for our V bar which was similar in many respects to the vise used by McEwen and Rodger (1960). Thus in Exp. II the I bar was 35 mm. wide and the T bar was 60 mm. wide. The V bar was a Stanley Woodworkers Vise No. 700 on either face of which had been affixed $\frac{1}{8}$-in. thick fiber boards. The movable face of the vise passed over a millimeter rule which permitted *E* to quickly and accurately set the V bar at a desired opening as well as record *S*'s PSE. The V bar settings prior to each judgment were at 45, 90, 60, 75, and 30 mm. After the V bar had been set and *S*'s hands placed on V bar and T bar, *E* turned the vise until *S* reported "equality." Other than these changes the procedure for Exp. II remained the same as that of Exp. I.

Results

Differences between groups on presatiation trials were not significant nor were interaction effects of Trials

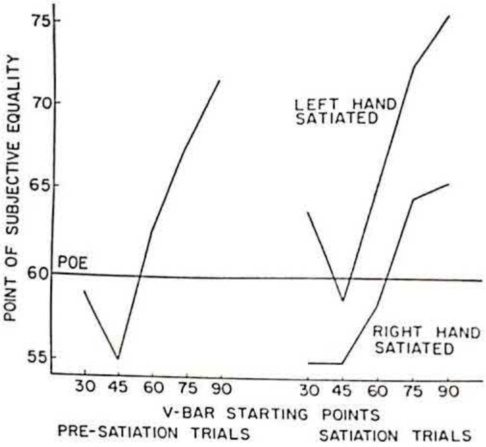


FIG. 2. Mean T bar width judgments before and after satiation (35-mm. I bar; 60-mm. T bar; V bar starting points in mm.).

× Groups. Again, as in Exp. I, the effect of starting point gave rise to significantly different judgments between trials, $F(4, 72) = 38.85$, $p < .01$.

Table 2 summarizes the results of an analysis of variance for satiation trials. The significant *F*s for groups and trials and the nonsignificant *F* for interaction effects suggests that the groups performed in the same fashion over trials but at different relative levels. It is possible that the greater distortion of judgments for the right-hand satiated *S*s seen in Exp. I was an artifact of the manipulanda.

Figure 2 graphically portrays the PSE curves for both presatiation and satiation trials. Again starting points are given in an increasing order rather than in the order of actual presentation to *S*. It is clear from Fig. 2 that the behavior reported in Exp. I is reliable. That is, left-hand satiated *S*s who went from the 35-mm. I bar to the 60-mm. T bar judged the T bar to be wider at every V bar setting than they did on presatiation trials. In contrast, right-hand satiated *S*s who proceeded from the 35-mm. I bar to the V bar, judged the 60-mm. T bar to be narrower in comparison to presatiation trials. Without attempting an explanation at this point, it is apparent that when the satiated hand proceeds from the I bar to the V bar the error in perceptual judgment is in favor of the I bar size, while the reverse is true for *S*s whose satiated hand proceeds from the I bar to the T bar.

EXPERIMENT III

Experiment III considered two further questions associated with KFAE generally and satiation theory in particular. Traditionally, satiation has been induced by having *S*s *rub* their fingers lightly over the I bar edges (e.g., Charles & Duncan, 1959; Heinemann, 1961; Krauskopf & Engen, 1960). While no explicit demand for this procedure appears in the literature it is presumed that such a practice was deemed advisable in order to create more readily a "locus of excitation" and a "field of electrotonic effects" from which T objects would recede. On the other hand, a contrast theorist would simply ask that *S*s form an "impression" or "judgment" of, in this case, the width of the I bar. They would then predict that *S*s' judgments of similar stimuli would be affected by the first "impression."

The second question relates to the notion of a "distance gradient" in KFAE (cf. Charles & Duncan, 1959). It is hypothesized that T objects placed at increasingly greater "distances" from the satiated boundaries of a previously inspected I bar will recede less, and finally not at all, from these boundaries. A contrast explanation would require no such assumption of "distance." Further, the Charles and Duncan (1959) manipulanda included a wedge-shaped V bar similar to that used in Exp. I of the present studies. While they interpreted the lack of distortion in judging a $\frac{1}{8}$-in. wide T bar to support the "distance gradient" hypothesis, it is clear that cues emanating from the end of the V bar may have been operating.

Method

The 24 *S*s in the present study were drawn from the same population as previously described. Procedures and manipulanda were precisely the same as in Exp. II on presatiation trials. After presatiation trials *S*s were assigned at random to one of four groups as follows: right hand on V bar after "no satiation"; right hand on V bar after satiation; left hand on T bar after "no satiation"; left hand on T bar after satiation. "No satiation" refers to a condition whereby *S*s merely placed thumb and fingers on either edge of the I bar without rubbing the edges in the traditional manner. In order to provide a somewhat greater contrast or "distance" between I and T bars than was provided in Exp. I and II, in this experiment the I bar was 110 mm. wide and the T bar, 35 mm.

Results

An analysis of variance of T bar judgments on presatiation trials revealed no significant differences between the groups nor were interaction effects of Groups × Trials significant. Significant differences appeared between trials in the same fashion as in the other two experiments, $F\ (4, 80) = 4.51$, $p < .01$.

TABLE 3

ANALYSIS OF VARIANCE OF RIGHT-HAND AND LEFT-HAND, "NO-SATIATION" AND SATIATION GROUPS' T BAR JUDGMENTS OVER FIVE SATIATION TRIALS

Source	*df*	*F*
Between *S*s		
Groups (B)	3	8.60**
Error (b)	20	
Within *S*s		
Trials (A)	4	26.34**
A × B	12	1.42
Error (w)	80	

** $p < .01$.

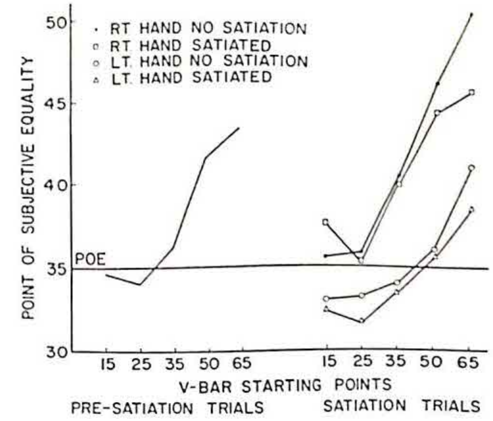


FIG. 3. Mean T bar width judgments before and after satiation (110-mm. I bar; 35-mm. T bar; V bar starting points in mm.).

Table 3 summarizes an analysis of variance of data obtained on satiation trials. The nonsignificant F for interaction effects of Groups × Trials coupled with the significant F for between-group variance, $F\ (3, 20) = 8.60$, $p < .01$, suggests that groups performed in a consistent fashion at their respective levels. The F for trials is again significant, $F\ (4, 80) = 26.34$, $p < .01$.

Figure 3 provides a clearer picture of the performance of the four groups on satiation trials. The single curve combining groups on presatiation trials is again provided for comparative purposes. Most apparent from Fig. 3 is that both left- and right-hand groups, whether satiated in the traditional manner or not, perform in a virtually identical fashion. It would appear as though the traditional satiation technique of rubbing the bar is unnecessary, and instructions to merely hold, but "appreciate" the bar's width, are sufficient. Further illustrated in Fig. 3 is the significantly different performance of right- and left-hand satiated groups in the directions expected on the basis of Exp. I and II.

Discussion

The satiation theory put forth by Köhler and Wallach (1944) was based primarily on visual aftereffects. The later Köhler and Dinnerstein (1947) study dealt with a satiated field explanation of KFAE. It is the latter application of satiation theory the present studies question.

The most apparent difficulty is that of the accounting for a single satiated area when *both* hands of *S* are involved in the critical judgment. Thus, Köhler and Dinnerstein caused *S*'s left hand to inspect an I bar after which this hand proceeded to a T bar. In this way they could and did hypothesize the existence of satiated boundaries of the I bar within which (usually) was projected the subsequently presented T bar. But what of the right hand? In the last analysis it is *S*'s right hand which indicates the POE on a variable bar. Inaccuracies of the right hand reveal the presence or absence of KFAE. The problem is that of accounting for the cortical locus of the right hand's information. Does it proceed to the same satiated area created by the left hand or to an entirely different location? In any event, it would seem as though a single cortical area is insufficient to account for the variety of sensory channels stimulated and that at some location a comparison or contrast between left- and right-hand stimuli is effected. The present studies sought to bring these and other questions to the fore.

Two aspects of these studies may be dispensed with rather briefly. First, *S*s who have been given the opportunity to form a perceptual impression of an object will judge a subsequent object in relation to the previous one. Simply holding a bar as in Exp. III is just as effective a means of altering subsequent perception as is the process of rubbing the bar.

Secondly, in Exp. III we cannot guarantee that the 35-mm. T bar was sufficiently "within" the satiated boundaries of the 110-mm. I bar so that satiation effects could be avoided. If satiation theory is correct then we must assume the 35-mm. bar was not far enough removed. On the other hand contrast theory would require no such assumption of distance. While Charles and Duncan (1959) employed I and T bar ratios more extreme than employed here and their data are in support of satiation theory, their use of a wedge-shaped V bar gives rise to questions of cues arising from the end of the V bar.

The behavior of *S*s whose left hands received the satiation experience, subsequently proceeding to the T bar, offers few problems of interpretation for either contrast or satiation theories. Satiation asserts that,

> T-objects recede from areas in which I-objects have been shown, particularly from regions in which the contours of these objects have been located [Köhler, 1951, p. 200].

Thus, in Exp. I the satiated boundaries of the 2.5-in. I bar caused the subsequently presented 1.5-in. T bar to "recede" perceptually and be judged narrower by *S*s who had previously judged the same bar more accurately. In Exp. III we have a similar situation and explanation with, however, a 110-mm. I bar and a 35-mm. T bar. In this case, despite the rather marked discrepancy in size, the boundaries of the 35-mm. T bar "receded" and this bar was judged to be narrower than 35 mm.

It is, of course, possible to interpret these data in terms of contrast. That is, after prolonged experience on the 2.5-in. or 110-mm. I bars, the 1.5-in. and 35-mm. T bars, in comparison, were judged narrower than they were when judged without the I bar experience. Similarly, a baseball bat feels lighter if a heavier bat is hefted first.

The same explanations pertain to Exp. II in which a 35-mm. I bar and a 60-mm. T bar were employed. In this case satiation theory would describe the T bar boundaries as falling beyond the previously satiated 35-mm. I bar boundaries. Even so, the spread of excitation from the I bar experience would cause the

T bar boundaries to recede, giving rise to a "wider" judgment than occurred without satiation experience. Contrast theory would again assert that the contrast from narrower to wide enhances the percept "wide." Again, a baseball bat feels "heavier" if one has become used to a light one.

It was apparent in the discussions of left-hand satiated *S*s that since Exp. I and III employed I and T bars which differed in the same direction, the explanation of the data was the same. Equally apparent was the fact that in Exp. II, in which the direction of the I and T bars' size difference was reversed, the explanation remains the same, but in a sense, reversed. Therefore to avoid repetition and confusion, the discussion from this point generally, and in relation to the right-hand satiated *S*s in particular, will refer only to the data of Exp. III.

The right hand of these *S*s after satiation experience proceeded to the V bar with their left hand to the T bar. The T bar under this procedure was judged wider than it was prior to satiation, and wider than comparable *S*s judged it to be under the left-hand satiated condition. The issue then, more pertinent to satiation theory than contrast, is that of accounting for the T bar's PSE being in a direction toward, rather than away, from the I bar boundaries. It seems necessary therefore, to consider two aspects of KFAE studies which have received relatively little attention in the past—namely, the role of "starting points" or V bar hand placement and the role of the unsatiated hand.

KFAE studies have generally counterbalanced starting points on the V bar as was done in the studies reported here. Unlike previous KFAE studies, however, the mean judgment (PSE) from *each* starting point is given in this series of studies rather than the more usual single measure. A close inspection of Fig. 3 will reveal the importance of the starting points.

After satiation on the 110-mm. I bar, *S*s of both groups had their right hand placed on the V bar which before each trial had been set at an opening of either 15, 25, 35, 50, or 65 mm. If we look at the behavior of both groups at each of these settings we find their judgments (i.e., their instructions to *E* to open or close the vise) proceed in the same direction for each setting save one. Hence with the vise set at either 15 or 25 mm., both groups move toward the T bar POE (35 mm.). The left-hand group underestimates; the right-hand group overestimates. Both groups are obviously moving toward, or to the boundary of, previously satiated tissue. Apparently, however, the satiated boundary differs according to the hand satiated. With the vise set at either 50 or 65 mm., both groups move away from satiated tissue and toward the POE, but again the boundary is not the same for both hands nor is it the same as it was for the 15- or 25-mm. V bar setting since the PSEs differ rather markedly. Only when the vise was set precisely at POE do the two groups move in opposite directions. The left-hand satiated group moved to a narrower setting; the right-hand group to a wider setting.

It is important now that we distinguish between our groups in a manner other than right and left hand satiated. We can distinguish between the right- and left-hand satiated groups if we refer to the former as an "indirect" contrast group and the latter as a "direct" contrast group. These terms simply note that the contrasting objects, the I and T bars, are felt by the same hand (sensing organ) in one case and by opposite hands in the other. Thus the left-hand satiated group makes a "direct" comparison (I bar to T bar) while the right-hand satiated group makes an "indirect" comparison (I bar to V bar).

At least three things seem clear from the present studies. Both satiation theory and contrast theory have difficulty explaining the "indirect" contrast (right-hand satiated) group. While contrast theory experiences no difficulty whatever in explaining the "direct" contrast (left-hand satiated) group, satiation theory does. It must resort to additional assumptions (e.g., the assumption of "distance" of T bar boundaries from I bar boundaries), and even then, this group does not perform in a consistent manner. That is, at times this group proceeds toward satiated areas and at other times away. In no case do the PSEs or "satiated boundaries" coincide. Secondly it is apparent that the PSEs are influenced not only by I and T bar sizes but by V bar settings. Thirdly, *Ss* provided with the opportunity to contrast stimuli directly are on the average more accurate in their judgments than *Ss* who must rely on an indirect comparison.

If a rule could be stated, it would be that when two objectively discrepant objects are directly compared any error manifested will be in the same direction as the objective discrepancy between the second object as compared to the first. Thus an I bar followed by a smaller T bar presented to the same hand, will find *Ss* judging the T bar to be smaller than it actually is. If two objectively discrepant objects are indirectly compared any error manifested will be in the same direction as the objective discrepancy between the first object as compared to the second. Thus an I bar presented to one hand, followed by a smaller T bar presented to the other hand, will cause *Ss* to judge the T bar to be larger than it actually is.

It would seem then, that the phenomenon we have been referring to as KFAE is but another aspect of that larger phenomenon involving man's responses to contrasting stimuli whether such stimuli be forces or pressures, auditory or visual, chemical or tactual. By focusing attention on external functions and manipulanda rather than on presumed internal states, it would also seem as though more accurate predictions covering a greater variety of stimulus conditions would result.

REFERENCES

CHARLES, J. P., & DUNCAN, C. P. The distance gradient in kinesthetic figural aftereffect. *J. exp. Psychol.*, 1959, **51**, 164–170.

DEMBER, W. *The psychology of perception.* New York: Holt, 1960.

GIBSON, J. J. Adaptation, after-effect, and contrast in the perception of curved lines. *J. exp. Psychol.*, 1933, **16**, 1–31.

HEINEMANN, E. G. Figural aftereffects in kinesthesis: Effects of object width and repeated presentations. *J. exp. Psychol.*, 1961, **61**, 51–56.

KÖHLER, W. Relational determination in perception. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951.

KÖHLER, W., & DINNERSTEIN, D. Figural aftereffects in kinesthesis. In, *Miscellania psychologia Albert Michotte.* Paris: Libraire Philosophique, 1947. Pp. 196–200.

KÖHLER, W., & WALLACH, H. Figural aftereffects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 263–357.

KRAUSKOPF, J., & ENGEN, T. The distance-paradox in KFAE. *Amer. J. Psychol.*, 1960, **73**, 298–301.

MCEWEN, P., & RODGER, R. S. Some individual differences in figural aftereffects. *Brit. J. Psychol.*, 1960, **51**, 1–8.

SPITZ, H. H. The Köhler-Wallach theory of satiation. *Psychol. Bull.*, 1958, **55**, 1–28.

(Received February 20, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 91–97

EFFECT OF OVERLAPPING CUES UPON DISCRIMINATION LEARNING[1]

CHARLES N. UHL[2]

University of Alberta

Predictions of the mathematical models of discrimination learning of Burke and Estes (1957), Estes (1959), and Restle (1957, 1958) were tested. Human Ss predicted which of 2 reinforcing events would appear in the presence of several stimuli each composed of 8 cues. Reinforcement was available 100% of the time with 1 set of stimuli (S_{100}) while only 50% reinforcement was available with a 2nd set of stimuli (S_{50}). The independent variable was the proportion of cues common to both sets of stimuli (C). Values of C were: 1, .75, .50, .25, and 0. For all values of C between 0 and .75, the probability of prediction of the reinforcing event in the presence of S_{100} increased at approximately the same rate to an asymptote at .98. These findings tend to favor the pattern model of Estes. The discrimination task used was found to be unsuited to a test of the component model of Burke and Estes.

A critical problem faced by mathematical models of discrimination learning is the status of irrelevant stimuli in the learning process. In particular, an adequate theoretical treatment of discrimination learning must account for the effects of overlapping stimuli whose appearance from trial to trial is uncorrelated with the appearance of the discriminative stimuli, e.g., background stimuli, response produced stimuli, and common components of the stimulus display. To assess the effects of overlapping stimuli, an experiment has been conducted in which *E* manipulated the proportion of overlapping stimuli appearing in the stimulus display. One set of nonoverlapping discriminative stimuli was associated with 100% reinforcement while a second set of stimuli was associated with reinforcement 50% of the time.

Before proceeding to a detailed consideration of the experiment and the predictions certain discrimination learning models make about its outcome, certain abbreviations and boundary conditions need to be introduced. Two sets of nonoverlapping discriminative stimulus elements are referred to as S_1 and S_2. A set of common (overlapping) elements is S_c. On a T_1 trial *E* samples elements from S_1 and S_c, and on a T_2 trial *E* samples elements from S_2 and S_c. The two mutually exclusive and exhaustive response choices are A_1 and A_2. The two mutually exclusive and exhaustive events which are predicted by a response are E_1 and E_2. The probability of E_1 on a T_1 trial is π_{11}, and π_{21} is the probability of E_1 on a T_2 trial. The probabilities of A_1 on T_1 and T_2 trials are referred to as P_{11} and P_{21}, respectively. Trials of Type T_1 and T_2 occur with equal frequency. The proportion of overlapping elements, C, in a sample of elements may take on any value between 0 and 1. A given value of C remains constant on all trials, both T_1 and T_2. Over trials, $\pi_{11} = 1$ and $\pi_{21} = .5$.

---

[1] Supported in part by Grant APA-79 from the National Research Council of Canada, Associated Committee on Experimental Psychology. The writer wishes to express his appreciation for the assistance of Frank Hawes and Norma Howard.

[2] Now at the University of Utah.

Three mathematical models of discrimination learning have been found to yield three clearly different predictions concerning the effects of overlapping stimulus elements (cues) upon the rate of change and asymptotic value of P_{11} and P_{21} in the situation described. Restle (1957, 1958) has presented a theory which assumes that the rate of discrimination learning depends upon the mean validity of all cues. Validity (V) of a cue is defined to be $(2\pi-1)^2$, where π is the probability of E_1, given the cue in question, averaged over T_1 and T_2 trials. Thus, $V(S_1) = 1$, $V(S_2) = 0$, and $V(S_c) = .25$ in the present situation. Restle assumes that the conditioning or adaptation of a cue takes place on a relational basis. A cue (a) will be conditioned if it has greater V than another cue (b) appearing at the same time, and Cue b will be adapted with respect to Cue a. However, a cue, b, adapted with respect to another cue, a, may be conditioned relative to a third cue, c, provided that $V_b > V_c$ and provided that the three cues, a, b, and c, never appear simultaneously. With respect to the situation under discussion, Restle's model predicts that S_c will become adapted relative to S_1 which will be conditioned, and S_2 will become adapted relative to S_c which will be conditioned. Since the rate of conditioning (θ) of a cue depends not only upon the difference between its V and the V of an alternative cue, but also depends upon the proportion of cues of a given amount of greater V,

$$\theta = \frac{n_a V_a - n_b V_b}{n_a + n_b}$$

where n is the number of cues, a or b. It follows that an increase in C reduces the rate of conditioning of S_1 cues on T_1 trials (consisting of $S_1 + S_c$ cues), and it increases the rate of conditioning of S_c cues on T_2 trials (consisting of $S_c + S_2$ cues). Given an increase in $C (0 \leqslant C < 1)$ in the experiment under discussion, Restle's model predicts that: (a) the rate of increase of P_{11} from trial to trial decreases; (b) the rate of increase of P_{21} increases; (c) $P_{11} = 1$ at asymptote; and (d) $P_{21} = .75$ at asymptote. If $C = 1$, the model predicts that $P_{11} = P_{21} = .75$ at asymptote.

Burke and Estes (1957), enlarging upon statistical learning theory, assume that the proportion of a set of cues conditioned to response A_1 will equal the probability of E_1 on all trials on which the cues occur; this proportion is referred to as ϕ here. Rate of conditioning of cues depends only upon the rate at which S samples all available cues, hence rate of conditioning is independent of C if C is not correlated with the number of cues available for sampling. The asymptotic probability of response A_1 is a weighted sum of ϕ's in which the weighting factor is the proportion of one set of cues of the total population of cues available for sampling on a given type of trial (T_1 or T_2). The asymptotic value of $P_{11} = C\phi_c + (1 - C)\phi_1$, and the asymptotic value of $P_{21} = C\phi_c + (1 - C)\phi_2$ in the situation under discussion, e.g., if $C = .5$, $P_{11,\infty} = .875$ and $P_{21,\infty} = .625$. It is apparent that the theories of Restle and of Burke and Estes are in direct disagreement concerning rates of conditioning and asymptotes of P_{11} and P_{21}.

Although Restle's model permits treatment of a pattern of cues as a single cue, its major emphasis is like that of the Burke and Estes model in which a discriminative stimulus is treated as a collection of independent elements. Such a view, as recognized by both theories, is an oversimplification of the properties of discriminative stimuli. In particular, the "elements" of a stimulus tend to form a stable structure which is duplicated more or less exactly on successive presentations of the stimulus in most discrimination learning situations. Estes (1959) has developed a pattern model of discrimination learning in which a pattern is assumed to be an indivisible stimulus unit, e.g., each pattern is unique although patterns may share elements. The probability of response A_1 to a pattern increases with an increase in the value of a conditioning parameter q to an asymptote equal to the probability of E_1 associated with the pattern. The value of q is fixed for any one experiment. When there is more than one pattern, the rate of increase of the probability of A_1 from trial to trial decreases with an increase in the number of patterns (N). Asymptotic probability of A_1 is equal to the probability of E_1 averaged over the patterns available for display (if each is displayed equally often). In the present experiment, in which N is a constant when $C < 1$, Estes' pattern theory predicts for all values of $C < 1$ that: (*a*) P_{11} increases at the same rate to an asymptote at $\pi_{11} = 1$, and (*b*) P_{21} remains fixed at a value of $\pi_{21} = .5$. If $C = 1$, P_{11} and P_{21} increase at an equal rate to an asymptote at $\pi_{11} = \pi_{21}/2 = .75$. Now, we have three theories in sharp disagreement. Given an increase in $C (0 \leqslant C < 1)$ in the situation under discussion, Burke and Estes predict no effect upon rate of learning but an effect upon asymptotic performance, Restle predicts no effect upon asymptotic performance but an effect upon rate of learning, and Estes predicts no effect upon either rate of learning or asymptotic performance. If $C = 1$, all three theories are in agreement in predicting $P_{11} = P_{21} = (\pi_{11}+\pi_{21})/2$ at asymptote.

The purpose of this research was to test the relative merits of the three theories cited by varying C in a discrimination situation designed to meet the restricting assumptions of all three theories insofar as possible. An adequate test requires that the discriminative stimuli be susceptible alternatively to component or to pattern perception. It also requires a minimization of the role of mediating response mechanisms by means of which human *S*s commonly partition a stimulus pattern along dimensions, e.g., mediating verbal responses to color, shape, etc., in a concept identification task. Such partitions would introduce a third class of stimulus units which would not be under the direct control of *E*. A situation in which $\pi_{11} = 1$ and $\pi_{21} = .5$ is best suited to the test since the value of P_{21} is a sensitive measure of the conditioning status of common cues.

METHOD

Subjects.—The *S*s were 120 undergraduates whose participation fulfilled a requirement of the introductory psychology classes from which they were drawn. They were assigned randomly to treatment conditions with the restriction that there be an equal number of *S*s in each condition.

Apparatus.—A vertical stimulus display panel 28 in. × 28 in. containing a 6 × 6 matrix of 1-in. pilot lights, 4 in. apart on centers, was located 8 ft. in front of *S*. A sloping front response panel was located on the table at which *S* sat. The response panel contained two button switches 6 in. apart and two 1-in. pilot lights, green on the left and red on the right, each 4½ in. above a switch. An opaque screen separated adjacent *S*s.

Experimental events were controlled by an automatic programing device. Separate counters recorded responses to each of the button switches for each *S*.

Design and procedure.—Values of *C* employed were 0, .25, .50, .75, and 1. Each stimulus consisted of the illumination of eight pilot lights in the display panel. An illuminated pilot light is identified with an

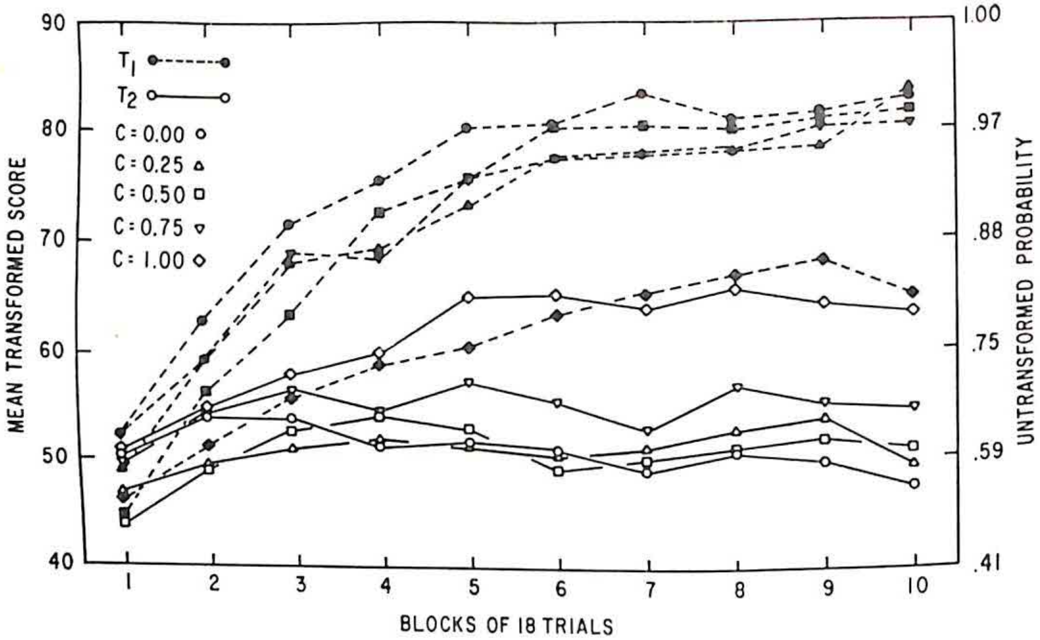


FIG. 1. Mean transformed P_{11} (T_1) and P_{21} (T_2) as a function of proportion of overlapping cues (C). (Untransformed probabilities of response appear on the right ordinate scale.)

element in the models of Restle and of Burke and Estes. The stimulus combination of eight pilot lights is identified with a pattern in Estes' model. Lights were randomly assigned to S_1, S_2, and S_c. A new assignment was made for each subgroup of *S*s. Once lights had been assigned to sets of elements, six random combinations of eight lights from $S_1 + S_c$ were formed for presentation as stimuli on T_1 trials, and six random combinations of eight lights from $S_2 + S_c$ were formed for presentation as stimuli on T_2 trials.[3] The appropriate value of C was maintained exactly in each stimulus, e.g., if $C = .25$, there were two S_c lights in each of the 12 stimuli. The probability of a light appearing in a stimulus combination was .5, and this probability was maintained exactly for all lights. The number of cues in S_1, S_2, and S_c consequently varied with the value of C, e.g., if $C = .25$, $n_1 = n_2 = 12$ and $n_c = 4$, but if $C = .75$, $n_1 = n_2 = 4$, and $n_c = 12$.

An experimental session consisted of 10 consecutive blocks at 36 trials each. The order of presentation of the 12 stimuli was random with the restriction that each must be presented three times in each block. Ten stimulus sequences, one per block, were prepared. Each session employed a random permutation of the 10 stimulus sequences. The left response button and event light were designated A_1 and E_1, respectively, for half of the *S*s in each subgroup, and the right button and light were A_1 and E_1, respectively, for the other half of the *S*s.

The *S*s were read instructions to the effect that they were to respond to the stimulus combination displayed on each trial by operating one of the two button switches on the response panel corresponding to which of the two event lights they expected to come on. Care was taken to make no reference to the stimulus combinations which would bias *S*s to respond to patterns or to component elements. The *S*s were urged to try to make as many correct predictions as possible. Temporal relations were as follows: stimulus duration, 3 sec.; interval from termination of the stimulus to onset of the event light, 1.5 sec.; duration of event light, 1 sec.; intertrial interval, 2.5 sec.

RESULTS

The data treated were the values of P_{11} and P_{21} taken over successive blocks of 18 trials. Due to a decrease in variance with an increase in P value, these data were subjected to an arc-sine transformation. Transformed *P*s comprised the data in all statistical analyses. Two analyses of variance of trends (Edwards, 1960) over 10 blocks of trials were performed separately for P_{11} and P_{21}.

---

[3] When $C = 1$, there were only six stimulus combinations.

Figure 1 shows curves of mean transformed P_{11} and P_{21} for each value of C over the course of training. It is apparent that the rate of increase and terminal value of P_{11} were approximately the same for all values of $C < 1$. The 90% confidence band of mean points of $C = 0.00$ included the mean points of all values of $C < 1$, and vice versa. Values of $C < 1$ had no appreciable effect upon the rate of increase and terminal value of P_{21}. The 90% confidence band of mean points of $C = 0.00$ included the mean points of all values of $C < 1$, and vice versa. There was a significant effect of values of C in analyses of variance of transformed P_{11} and P_{21}, $F\ (4, 115) = 10.83$ and 10.15, respectively, $p < .001$. The $C \times$ Blocks of Trials interaction was significant for both transformed P_{11} and P_{21}, $F\ (36, 1035) = 1.70$ and 1.89, respectively, $p < .01$. However, when the $C = 1$ data were eliminated in two supplementary analyses of variance, the only significant effect at the .10 level was blocks of trials, $F\ (9, 828) = 111.57$ and 3.76 for P_{11} and P_{21}, respectively, $p < .001$.

Separate analyses of variance of P_{11} and P_{21} trends of transformed P_{11} and P_{21} were performed on the last four blocks of trials. Neither effect of blocks of trials was significant at the .10 level, indicating that both P_{11} and P_{21} were approaching (at) asymptotic values. In fact, the curves of P_{21} for all values of $C < 1$ appear to be declining after the third block of trials. However, a nonsignificant quadratic orthogonal polynomial component of blocks of trials fails to lend support to the reliability of the latter trend at the .05 level. When $C < 1$, $P_{11} = .98$, untransformed, at asymptote. When untransformed asymptotic values of P_{21} are examined, $P_{21} = .59$ when $C < 1$. When $C = 1$, $P_{11} = .81$ and $P_{21} = .79$, untransformed, at asymptote. Neither of these latter P values differs significantly by t test at the. 10 level from the value of .75 predicted by all three theories under examination.

Discussion

Among the three theories of discrimination learning whose predictions this research was designed to test, the pattern model of Estes (1959) fared best when confronted by the data. Considering P_{11} values first, both Estes (1959) and Restle (1957, 1958) predicted asymptotic values of unity when $C < 1$, while Burke and Estes' (1957) component model predicted asymptotes decreasing from unity to .75 as C increases from zero to unity. Both Estes' and Restle's predictions were in close agreement with the data, while Burke and Estes' predictions were underestimates except in the case where $C = 1$. Restle's prediction that the rate of increase of P_{11} would decrease with an increase in C ($0 < C < 1$) did not receive statistical support.[4] However, a nonsignificant tendency was observed for the rate of increase of P_{11} to be most rapid when $C = 0$. There is a possibility that a more powerful experiment might have supported Restle's predictions concerning rate of increase of P_{11} as a function of C. Estes' pattern model successfully predicted the apparent independence of rate of learning as a function of $C < 1$. In the case of P_{21}, Estes' pattern model predicted an asymptote at .5, Restle predicted .75, and Burke and Estes predicted values decreasing from .75 to .5 as C decreases from unity to zero. Estes' prediction was closest to the observed asymptote at .59 when $C < 1$. There was no indication that asymptotic values of P_{21} increase with an increase in C when $C < 1$ as predicted by Burke and Estes. Statistical support was lacking for Restle's prediction that the rate of increase of P_{21} increases with higher values of C. Although Restle predicts a slow rate of increase of P_{21}, there was no indication that P_{21} was increasing at all beyond the fourth block of trials when $C < 1$. All three theories successfully predicted the results when $C = 1$, which was a case of nondiscriminative probability learning similar to that of Estes and Straughan (1954) and yielding the familiar probability-matching phenomenon.

---

[4] However, failure to reject the null hypothesis of equal rates does not prove that hypothesis. See Grant (1962) for a discussion of this problem in tests of quantitative models.

There can be little question that *S*'s behavior was most consistent with a pattern model. Undoubtedly, the stable structure of cues comprising a stimulus was more conducive to pattern identification than to identification of discrete cues. To this extent the nature of the stimuli biased the test against the Burke and Estes component model. However, the fact remains that *S*s chose to respond to pattern rather than to components of those patterns. Estes and Hopkins (1961) found that *S*s approached perfect discrimination of patterns made up of constituent subpatterns which were separately associated with different events. The conclusion seems inescapable that *S*s generally will attach responses to patterns whenever there is a set of recurrent stimulus structures reliably associated with a reinforcing event. The range of conditions in which this result is obtained has yet to be determined, however. Estes and Hopkins' results support the present finding that the rate of discrimination was unaffected by the value of C, if $C < 1$. As long as two stimulus structures are discriminably different, the extent of overlapping of components seems to have no effect upon pattern discrimination. However, this generalization may not apply in cases where there is a high degree of similarity of the structures of stimulus patterns, e.g., two or more ellipsoidal figures. Also, the minimum degree of discriminability remains to be ascertained.

The present findings are in substantial agreement with Part II of the study of Estes and Burke (1955) in which only background stimuli constituted S_c. However, Estes and Burke emphasized the results of Part I of their experiment, in which P_{11} failed to approach unity, as being due to the effects of S_c cues. Part II of their study plus the present findings, both based on a larger number of trials, suggest that P_{11} eventually will reach unity if $C < 1$. The present results indicate that the rate at which P_{11} approaches unity will not vary if $C < 1$. The present data resemble the tendency in Estes and Burke's data for P_{21} to increase above .50 early in training and then to drift back toward .50. Support is found for their suggestion that the conditioning of responses to S_c cues is reduced during the course of training. The manner in which such a reduction occurs in the weight of S_c remains unclear, although Restle's (1955, 1957, 1958) adaptation model offers one possible explanation provided that the adaptation of S_c cues on T_1 trials transfers to T_2 trials.

The failure of P_{21} to equal .50 at the end of training may have been due either to an insufficient number of trials or to a stable level of conditioning of S_c elements. The latter interpretation would mean that S_c has no effect upon probability of response when 100% reinforcement of a stimulus is available, but that it does have an effect if 100% reinforcement is not available. Verification of this hypothesis must await further experimentation.

REFERENCES

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, 22, 133–145.

EDWARDS, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Rinehart, 1960.

ESTES, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford, Calif.: Stanford Univer. Press, 1959. Pp. 9–52.

ESTES, W. K., & BURKE, C. J. Application of a statistical model to simple discrimination learning in human subjects. *J. exp. Psychol.*, 1955, **50**, 81–88.

ESTES, W. K., & HOPKINS, B. L. Acquisition and transfer in pattern vs. component discrimination learning. *J. exp. Psychol.*, 1961, **61**, 322–328.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GRANT, D. A. Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychol. Rev.*, 1962, **69**, 54–61.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

RESTLE, F. Theory of selective learning with probable reinforcements. *Psychol. Rev.*, 1957, **64**, 182–191.

RESTLE, F. Toward a quantitative description of learning set data. *Psychol. Rev.*, 1958, **65**, 77–91.

(Received February 21, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 98–99

SUPPLEMENTARY REPORTS

VARIATION OF POSTFEEDBACK INTERVAL IN SIMPLE MOTOR LEARNING

DOROTHE R. WEINBERG, DONALD E. GUY, AND RONALD W. TUPPER

University of Utah

The effects of varying postinformative feedback (post-IF) intervals (1, 5, 10, and 20 sec.) on learning a simple motor task were investigated. It was concluded that increasing this interval facilitates performance up to an optimum (5 sec.). An interpretation was suggested which provides a resolution of conflicting results found in earlier experiments using varied intertrial and post-IF intervals in widely different tasks.

In recent studies Bilodeau and Bilodeau (1958) and Denny, Allard, Hall, and Rokeach (1960) found that increasing the time interval between presentation of informative feedback (IF) on one trial and the occurrence of the "ready" signal for the next retards performance in simple motor tasks, e.g., learning to pull a yardstick 10 in. from a sheath while blindfolded. According to Denny et al. this deterioration is particularly noticeable when the response to be learned is unidentified, i.e., when IF is given in unspecified "units" of error.

With a more complex task, concept identification, Bourne and Bunderson (1963) also found post-IF interval to be a crucial factor affecting performance. However, in this study, increases in post-IF interval produced uniform linear *decreases* in mean number of errors. In addition to obvious task differences, the motor skills and concept identification experiments explored essentially nonoverlapping ranges of post-IF intervals. The shortest interval used by Bilodeau and Bilodeau was 8 sec., while the longest interval of Bourne and Bunderson was 9 sec. Discrepant results of these experiments may be resolved by an investigation of short post-IF intervals, on the order of those used by Bourne and Bunderson, in a motor skills task. On the assumption of a singular fundamental effect, it is hypothesized that increases in the post-IF interval in a motor skills task will improve performance up to some optimum beyond which the decremental effect of Bilodeau and Bilodeau will obtain.

Method.—The apparatus was a duplicate of Bilodeau and Bilodeau's stick-pulling task used in their Exp. 5. The *S* was to learn to pull a yardstick 10 in. out of a sheath while blindfolded. After *S*'s response on each trial, *E* stated a number (IF) which indicated the extent of error, above or below 10 in., in unidentified units (actually $\frac{1}{8}$-in. units). All *Ss* were given 25 trials, each of which began with a "ready" signal, followed in order by *S*'s response (5 sec. allowed maximum), IF (immediate verbal description of the extent and direction of *S*'s error), and the post-IF interval (during which *S* rested his hand). Forty *Ss*, enlisted voluntarily from introductory psychology classes, were assigned in equal number to four groups differing in length of post-IF interval, which took values of 1, 5, 10, and 20 sec.

Results and discussion.—An analysis of variance performed on extent of absolute error showed no overall significant variation among the four post-IF intervals in stick-pulling performance. However, response accuracy, which was nearly asymptotic after 25 trials, reached markedly different terminal levels. The interaction of trials by post-IF intervals was significant, $F\ (12, 144) = 3.01$, $p < .01$. A further analysis of variance, performed on the final 10 trials only, showed significant variation among the four post-IF intervals, $F\ (3, 36) = 5.40$, $p < .01$. A contrast of the 1-sec. group with the remaining three groups was highly significant, $F\ (1, 36) = 7.99$, $p < .01$, accounting for 87% of all variance attributable to post-IF conditions. Thus, the 1-sec. group produced reliably worse performance than the other three groups. No other contrast approached an acceptable significance level.

The results of this study extend and help to unify the experimental findings on the effect of variation in the post-IF interval in two grossly different tasks, motor learning and concept identification. They are strictly comparable to the results of Bourne and Bunderson, who concluded that the post-IF interval is the locus of at least some intratrial learning. As the length of the interval increases, so does the opportunity for *S* to make use of the information provided on any trial, thus facilitating overall performance. A reasonable interpretation of the conjoined results of this experiment and the Bilodeau and Bilodeau study is that intratrial learning in simple motor tasks is effectively complete after some short time (approximately 5 sec.) and that with longer intervals, as Bilodeau and Bilodeau suggest, some forgetting may occur. Altogether, available data suggest an optimum post-IF interval resulting from simultaneous acquisition and forgetting effects. This prediction should be further explored in a study of concept identification using a broader range of post-IF intervals than that of Bourne and Bunderson.

REFERENCES

BILODEAU, E. A., & BILODEAU, I. McD. Variation of temporal intervals among critical events in five studies of knowledge of results. *J. exp. Psychol.*, 1958, **55**, 603–612.

BOURNE, L. E., JR., & BUNDERSON, C. V. Effects of delay of informative feedback and length of post-feedback interval on concept identification. *J. exp. Psychol.*, 1963, **65**, 1–5.

DENNY, M. R., ALLARD, M., HALL, E., & ROKEACH, M. Supplementary report: Delay of knowledge of results, knowledge of task, and intertrial interval. *J. exp. Psychol.*, 1960, **60**, 327.

(Received January 21, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 99–101

EFFECTS OF PROMPTING AND CONFIRMATION IN A SERIAL LEARNING TASK[1]

JAMES R. HAWKER[2]
Washburn University of Topeka

60 Ss learned an electrical punchboard maze with half the Ss following a prompting procedure and the others following a confirmation procedure. A prompting procedure is one in which S is prohibited from making incorrect responses while confirmation refers to trial-and-error learning. The results indicated no differences in the 2 procedures in terms of number of errors or trials to reach criterion. However, the results did indicate that confirmation resulted in a better final performance as measured by time required to traverse the maze. While the present study failed to support the finding from paired-associate studies that prompting leads to better performance than confirmation, the results do support those from "response-guidance" studies of maze learning.

Several studies (Angell & Lumsdaine, 1960; Cook, 1958; Cook & Spitzer, 1960) have indicated that a "prompting" training procedure is more effective in the learning of paired-associate material than is a "confirmation" procedure. A prompting procedure is one in which *S* is prohibited from making errors (i.e., he is prompted on the correct response), while confirmation refers to trial-and-error learning. Since all of these studies have used the same paired-associate materials, the present study was designed to test the relative effects of the two procedures in another learning task—maze learning. Thus, the results should indicate the generality of previous research findings.

Method.—The apparatus was an electrical punchboard maze which contained 60 metal discs (10 rows × 6 columns) with a small pilot light mounted above each disc. Electrical power was furnished by a 6-v. battery, and the maze contained two independent circuits so that the discs and lights could be connected in two different ways. One circuit permitted a prompting technique to be used, while the other permitted confirmation. A series of switches enabled *E* to use either or neither of the circuits. The entire apparatus was enclosed in a wooden box with the face of the maze set at a 45° angle so as to be parallel to *S*'s face. An electric counter recorded the number of incorrect responses, and was enclosed in a separate, insulated box so that *S* could not hear it.

The *Ss* were 60 male undergraduate college students who volunteered for the experiment.

---

[1] This research was carried out at North Carolina State College, Raleigh.

The author wishes to thank John O. Cook for his advice and assistance.

[2] Now at the University of Pittsburgh.

TABLE 1

NUMBER OF ERRORS AND TIME TO TRAVERSE MAZE IN BLOCKS OF TWO TEST TRIALS

Measure	Group	Trials				
		1–2	3–4	5–6	7–8	9–10
Errors	Prompting					
	M	9.73	5.20	2.53	1.73	1.77
	SD	3.42	4.45	3.04	2.89	3.21
	Confirmation					
	M	10.73	5.17	2.80	.77	.60
	SD	3.29	3.85	3.52	1.43	1.90
Time	Prompting					
	M	24.98	21.15	17.25	15.45	14.23
	SD	10.63	8.03	7.45	6.58	5.64
	Confirmation					
	M	26.08	21.77	16.33	12.85	10.80
	SD	9.90	9.37	9.05	5.80	3.76

The *S*s were randomly assigned to one of two groups upon appearing at the laboratory with the restriction that there should be an equal number of *S*s in each group. Each *S* was run individually.

One group, Group C, followed a confirmation procedure. In this procedure, when *S* touched the correct disc in each row with a metal stylus, the light directly over it came on confirming his response as correct. The other group, Group P, followed a prompting procedure. In this procedure, when *S* touched the "Start" disc, the light over the correct disc in the *first* row came on; then, when *S* touched that disc, the light over the correct disc in the *next* row came on, and so on. Thus, *S* was shown the correct response at each point.

Since the prompting procedure entailed error-free responding, it was necessary to interpolate test trials during the practice trials so that a measure of learning could be obtained. Thus, each series consisted of two practice trials following the particular training procedure and a test trial during which *S* traced the pathway through the maze without the benefit of the lights. Consequently, the test trial was identical for both groups. All *S*s were given 10 series of trials and, if they had not reached a criterion of two consecutive errorless test trials by that time, were continued until they did reach the criterion.

Results and discussion.—During each test trial, the amount of time (in seconds) required to traverse the pathway and the number of incorrect responses were obtained. Analysis of variance was performed on both sets of data, and neither analysis indicated any overall differences between the two training procedures, $F\ (1, 58) < 1.00$ in both analyses.

The data were combined into blocks of two test trials each (see Table 1), and a repeated-measures analysis was carried out on both sets of data to see if there was a differential effect of training procedure over trials. The analysis of the error scores showed significant effects, of course, due to trials, but the Trials × Training Procedure interaction (T × P) failed to reach significance, $F\ (4, 232) = 1.90$. In the analysis of the time scores, there were again significant effects due to trials; moreover, the analysis indicated a significant T × P interaction, $F\ (4, 232) = 2.60$, $p < .05$. The interaction indicated that the prompting procedure resulted in faster "running speeds" during the initial part of training, but that confirmation resulted in a faster final running speed.

Since all *S*s were carried to a mastery criterion, an analysis was made of the total number of trials required to reach the criterion. The mean numbers of trials for Groups P and C were 17.40 and 16.50, respectively; again, analysis of variance indicated no differences between the two groups, $F\ (1, 58) < 1.00$.

The present study does not support the finding from verbal learning studies that prompting is a more effective training procedure than confirmation. There were no differences between the two groups in number of errors, or in number of trials to reach a mastery criterion. In fact, the data indicated that confirmation results in a better final performance as measured by time required to traverse the maze.

While the present results failed to agree with those from paired-associate studies, they are in accord with the results obtained in the response-guidance (cf. prompting) studies of maze learning of several decades ago. Carr (1930), in summarizing a number of these studies, indicated that response guidance had a sporadic effect on learning, and that its effectiveness seemed to depend upon the particular task involved. Moreover, he noted that response guidance led to a slower final speed in tracing a maze.

Consequently, it appears that the prompting and confirmation procedures need to be investigated more thoroughly in a number of different learning situations.

REFERENCES

Angell, D., & Lumsdaine, A. A. Prompted plus unprompted trials versus prompted trials alone in paired-associate learning. *USAF Off. Scient. Res. tech. Note*, 1960, No. 60-808.

Carr, H. Teaching and learning. *J. genet. Psychol.*, 1930, **37**, 189–217.

Cook, J. O. Supplementary report: Processes underlying learning a single paired-associate item. *J. exp. Psychol.*, 1958, **56**, 455.

Cook, J. O., & Spitzer, M. E. Supplementary report: Prompting versus confirmation in paired-associate learning. *J. exp. Psychol.*, 1960, **59**, 275–276.

(Received January 29, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 1, 101–102

CUE SELECTION AND MEDIATED TRANSFER IN PAIRED-ASSOCIATE LEARNING[1]

JAMES J. JENKINS and VIRGINIA B. BAILEY
University of Minnesota

A replication was performed of the Underwood, Ham, and Ekstrand experiment on cue selection in paired-associate learning. The portion of their experiment dealing with trigram-color stimulus compounds was exactly replicated. Original learning used compound stimuli, transfer was tested to each stimulus component alone. In addition 1 group of Ss was transferred to appropriate color names and ½ the Ss were required to spell the trigram stimuli aloud during learning. Ss were 80 undergraduates. Transfer was virtually complete to colors and color names but very poor to trigrams. Spelling the trigrams aloud had no statistically significant effect on the transfer performances.

Underwood, Ham, and Ekstrand (1962) showed that when *S* is presented a compound stimulus in paired-associate learning, cue selection may occur. They used seven word-color or trigram-color compound stimuli with the numerals 2 through 8 as responses, followed by a transfer test in which one or the other component was presented alone as the stimulus. A control group continued with the original compound. In the case of the trigram-color compound, complete transfer was found to color as a stimulus but only small positive transfer for the trigrams.

Method.—The present experiment repeated the trigram-color portion of the experiment with two additional variations. First, half of the *S*s were required to spell aloud the trigram portion of the stimulus during initial learning and the relevant transfer condition. (This procedure was designed to eliminate the possibility that *S*s simply ignored the more difficult portion of the stimulus.) Second, a new transfer condition was added. One group was transferred to printed word stimuli which named the relevant color portion of the original compound stimulus (e.g., RED, BLUE, YELLOW, etc.).

If, as many mediational theorists suggest, *S*s use verbal labels during initial learning, or, if the mediational responses to colors and color names have high overlap, or, if *S*s have

TABLE 1

Mean Number of Trials to Criterion in Original Learning and Mean Number of Correct Anticipations in Transfer

Trigrams	Original Learning Trials		Transfer Anticipations	
	M	*SD*	*M*	*SD*
Not spelled				
CON	16.8	9.6	67.4	2.5
C	15.6	5.2	64.6	4.5
CW	10.8	5.3	66.9	2.2
T	12.1	4.3	35.1	11.3
Spelled				
CON	13.3	3.7	68.0	1.8
C	18.9	7.1	64.5	5.3
CW	16.6	6.5	64.8	5.0
T	15.1	4.7	42.0	10.0

---

[1] This research was supported by a grant from the National Science Foundation. We are grateful to B. J. Underwood for supplying details of the original study.

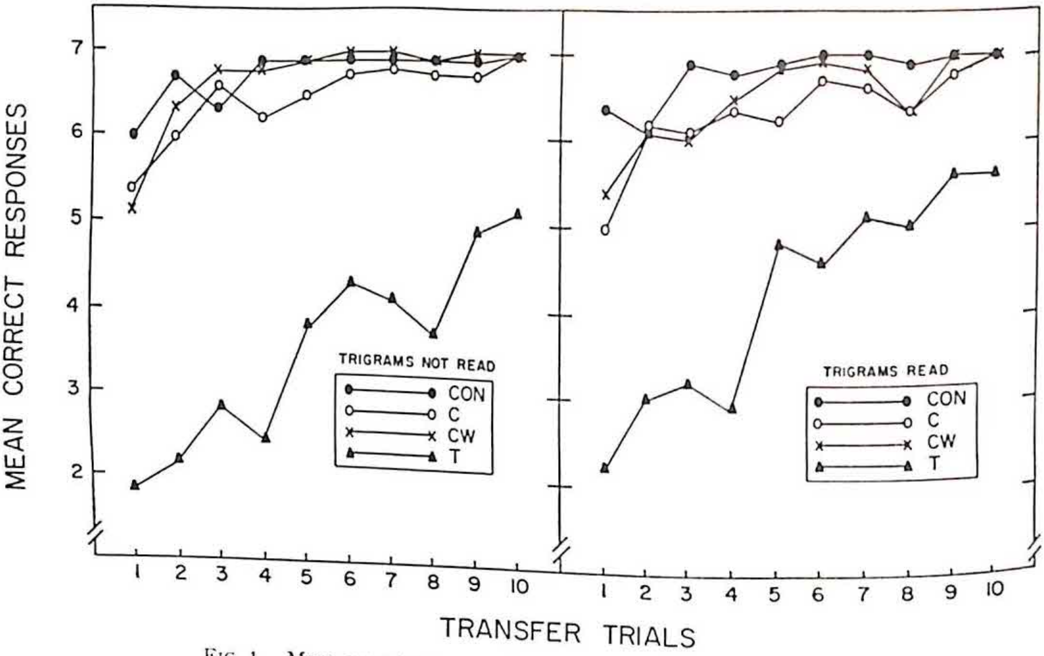


FIG. 1. Mean correct responses for each group on 10 transfer trials.

very strong associative links between colors and color names, the transfer to the color name as stimulus should be virtually as efficient as transfer to the color itself.

With these two exceptions, the material used and procedures followed were exactly the same as those of Underwood et al. (1962).

There were four transfer conditions: control (CON) with the original compound stimulus, trigrams alone (T), colors alone (C), and color words alone (CW). Half the *Ss* in each group were required to spell the trigram stimulus aloud in initial learning. There were thus eight conditions. There were 10 undergraduate *Ss* in each condition.

Results.—Table 1 shows the mean number of trials to learn the original list to one perfect recitation and the mean number of items correctly anticipated in 10 transfer trials for each group. A 2 × 4 analysis of variance on the original trials revealed no significant variation from group to group.

A 2 × 4 analysis of variance of the transfer scores showed a significant effect for transfer conditions, $F\ (3, 72) = 87.3$, $p < .001$, but no effect for reading the stimulus or for interaction between transfer conditions and reading.

The mean performance of each transfer group on each trial is shown in Fig. 1. The left-hand section refers to groups which did not spell the trigrams aloud, the right-hand section to those which did. It is readily apparent that no reliable differences are found between the group which continued with the original stimulus compound, the color transfer group and the color-name transfer group in either condition. The highly significant difference between the transfer conditions is clearly attributable to the poor performance of the groups transferred to the trigrams alone.

Spelling the trigram stimulus aloud during learning had no significant effect on the performance of the transfer task. The *Ss* who spelled trigrams aloud transferred to colors and color names as readily and to trigrams as poorly as groups which were not required to read the letters. While the trigram group which had spelled the stimuli made more correct responses on transfer than the trigram group which had not, the performances were not significantly different, $t\ (18) = 1.37$, $p > .10$.

When the data are pooled without respect to pronouncing the stimulus, they completely confirm the findings of Underwood, Ham, and Ekstrand.

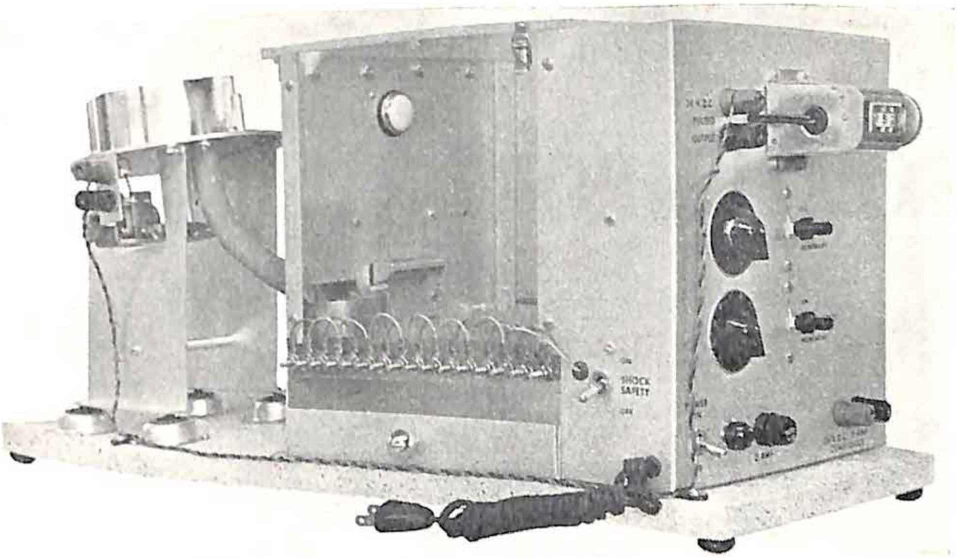
With respect to this specific stimulus compound, it may be concluded that (*a*) color is a completely effective stimulus; (*b*) colors may be replaced by color words without loss, indicating very effective mediated transfer; and (*c*) directing *S*'s attention to the more difficult portion of the stimulus compound by requiring him to read the trigram aloud has no significant effect on the functioning of this component in transfer.

REFERENCE

UNDERWOOD, B. J., HAM, M., & EKSTRAND, B. Cue selection in paired-associate learning. *J. exp. Psychol.*, 1962, 64, 405–409.

(Received February 19, 1963)

NEW MODEL A-105



Student Lab or Demonstration Box Animal Cage with Many Features by Scientific Prototype

This new Scientific Prototype animal cage combines cage, pellet feeder, power supply, light and shock source intensity controls, and counter. The unit is completely wired, mounted on a base, and ready to operate.

The rat box is approximately 9 × 8⅝ × 8. Grids are stainless steel. Top and sides are ¼" clear Plexiglass. The stainless steel pedal lever is counterbalanced to permit sensitivity adjustments. Dropping tray is aluminum anodized.

Of special interest is the feature of self-contained cue light and shock source *intensity controls.*

Price . . . $249.50 F.O.B. N.Y.C. . . . quantity discounts. Full information available.

Behavior Analysis Apparatus . . . operant conditioning kits, tachistoscopes, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, timers, switches, digital, modular test equipment; and special systems

SCIENTIFIC PROTOTYPE MFG., CORP.
623 W. 129th STREET • NEW YORK, N. Y.
TELEPHONES: (212) MO 6-4674 and 4789

the monthly review journal
of the
American Psychological Association

CONTEMPORARY PSYCHOLOGY

reviews

criticism

opinion

on

the latest books, films,

and other instructional media

lists of new books letters-to-the-editor

$1.00 a copy $10.00 a year
($10.50 foreign)

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333-16th Street, N.W.
Washington, D.C. 20036

Journal of
Experimental Psychology

Vol. 67, No. 2 February 1964

INFORMATION CAPACITY OF DISCRETE MOTOR RESPONSES[1]

PAUL M. FITTS AND JAMES R. PETERSON[2]

University of Michigan

The effects of response amplitude and terminal accuracy on 2-choice reaction time (RT) and on movement time (MT) were studied. Both the required amplitude (A) of a movement, and the width (W) of the target that S was required to hit, had a large and systematic effect on MT, whereas they had a relatively small effect on RT. Defining an index of movement difficulty as ID = $\log_2 2A/W$, the correlation between ID and MT was found to be above .99 over the ID range from 2.6 to 7.6 bits per response. Thus the times for discrete movements follow the same type of law as was found earlier to hold for serial responses. The relative independence of RT and MT is interpreted as pointing to the serial and independent nature of perceptual and motor processes.

The relation of response variability and response magnitude has long been one of the major topics of interest in psychophysics. Recently the effects of these two variables plus that of response duration have received renewed emphasis in connection with studies of the information capacity of the human motor system.

For physical communication systems having limited bandwidth (W) and signal power (S), and perturbed by white Gaussian noise of average power (N), Shannon (1948, Theorem 17) showed that channel capacity (C) is equal to

$$C = W \log_2 \frac{S + N}{N} \quad \text{bits per second} \qquad [1]$$

Arguing by analogy, Fitts (1954) reasoned that the average amplitude (A) of a human movement is equivalent to average signal plus noise amplitude, and that half the range of movement variability is equivalent to peak noise amplitude (n), and proposed that an index of task difficulty (ID) be defined as

$$\text{ID} = \log_2 \frac{A}{n} \qquad [2]$$

In an experiment in which the permissible plus-and-minus range of terminal movement error is specified as a target band of width (W), and S is induced to keep the proportion of his responses exceeding this limit small, it can be assumed further that $n = W/2$ and therefore ID can be evaluated as

---

[1] This research was supported in part by the Air Force Office of Scientific Research, Contracts No. AF 49 (638)-449 and -1235.

[2] Now at Minneapolis-Honeywell Company, Minneapolis.

$$\mathrm{ID} = \log_2 \frac{2A}{W} \qquad [3]$$

On these grounds it was conjectured that average movement time (MT) should remain constant for different values of A and W, within limits, as long as the ID ratio remains constant, i.e.,

$$\frac{\mathrm{ID}}{\mathrm{MT}} = C \qquad [4]$$

where C may be interpreted as analogous to man's capacity for executing a particular class of motor responses in bits per second.[3] Extending the analogy to the case where ID takes on different values, then if motor processes follow the same type of law as do perceptual-motor processes, it would be expected that, within limits,

$$\mathrm{MT} = a + b\,\mathrm{ID} \qquad [5]$$

Fitts (1954) obtained results in agreement with these predictions, using three different self-paced, cyclical tasks. Performance was maximum for a midrange of values of A, W, and ID and fell off gradually outside these values (optimum performance being obtained in the vicinity of $A = 8$ in. and ID = 4–8 bits, depending on the particular task). Subsequent investigators (Annett, Golby, & Kay, 1958; Crossman, 1960; Welford, 1958) have, with minor exceptions, reported findings that provide general confirmation of the theory. The times required for different components of a movement sequence appear to follow somewhat different rules, however, and a more complex relation between A and W has been found in certain tasks (Andriessen, 1960; Vredenbregt, 1959). Welford (1960, 1961) has also proposed the adoption of a modified ID ratio. However, all of these results have been obtained in repetitive self-paced tasks, where the only uncertainty confronting S in regard to the requirements of the next response in a series has been the uncertainty resulting from his own amplitude variability in executing the previous response. It therefore was decided to conduct further studies of the generality of the theory when applied to discrete movements. Separate measures of reaction time (RT) and movement time (MT) were taken, and variations introduced in the degree of uncertainty as to what response would be required by the next stimulus.

The experiments were planned to provide answers to the following specific questions: (*a*) Does a single index of performance apply equally well to both discrete and serial tasks? (*b*) Are RT and MT affected in the same way by variations in task conditions?

METHOD

Apparatus and task.—The *S*s held a lightweight metal stylus on a small metal starting plate (see Fig. 1), fixated a point midway between two signal lights, and when one of the lights came on attempted to hit the appropriate one of two alternative targets as quickly as possible. Target widths were varied (all were 4 in. long); targets could be located at any desired distance to right or left of the starting point.

An electrical current, sufficient in amount to activate an electronic relay, but below the threshold for detection, passed through *S*'s body whenever the stylus was in contact

---

[3] The analogy with Shannon's Theorem 17 is not exact, since the ID ratio involves an amplitude rather than a power ratio, and since the range of variability is used in estimating noise amplitude.

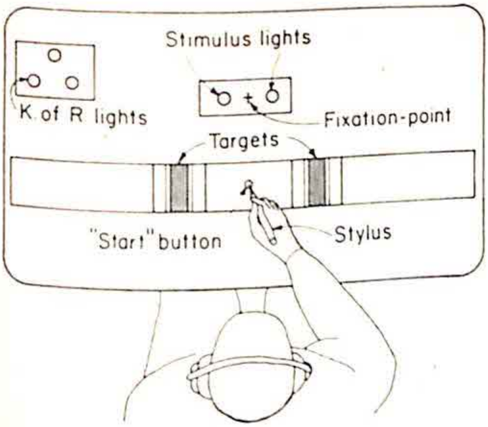


FIG. 1. Plan view of the apparatus.

TABLE 1

TARGET CHARACTERISTICS EMPLOYED IN THE SERIES OF EXPERIMENTS

Target	Movement Amplitude (*A*, In.)	Target Width (*W*, In.)	Index of Difficulty (ID, Bits)
A	3	1.0	2.58
B	3	0.5	3.58
C	3	0.25	4.58
D	3	0.125	5.58
E	6	1.0	3.58
F	6	0.5	4.58
G	6	0.25	5.58
H	6	0.125	6.58
I	12	1.0	4.58
J	12	0.5	5.58
K	12	0.25	6.58
L	12	0.125	7.58

either with the starting button, the target plate, or overshoot and undershoot plates on either side of the target. Three time clocks were used to measure RT, MT, and total time to the nearest .01 sec. Lights for knowledge of results provided immediate information regarding hits, overshoot errors, and undershoot errors.

Subjects.—Six paid male *S*s, 18–25 yr. of age, were used.

Conditions of A *and* W.—Twelve combinations of movement amplitude (*A*) and target width (*W*) were provided by using 3-, 6-, and 12-in. values for *A*, the distances between the starting button and the midline of the target, in combinations with .125-, .25-, .5-, and 1-in. values for *W* (see Table 1).

Procedure.—A series of five studies was carried out (see Table 2). The conditions, including target or stimulus probabilities, were explained fully to *S*s. The *S*s were instructed to make quick movements aimed at hitting one of two alternative targets as soon as one of the two stimulus lights appeared. Five trials were given per minute, and about 1 min. was required to change targets within blocks of trials. The order of target conditions within experiments was randomized and different for each *S*. The same six *S*s served in all five experiments following a training session.

The basic condition was that used in the Practice Session and in Exp. II (see Table 2). Two identical targets were located to the right and left of the starting button, and the right and left stimulus lights were turned on with equal frequency in random order. This task is thus a standard two-choice *B* reaction. In Exp. I one of each pair of targets, sometimes the right and sometimes the left one, was designated in advance but *S* was told that the stimulus indicating he should initiate a movement to that target would follow the ready signal only 50% of the time, and that no stimulus would occur on the other half of the trials. This is similar to an *A* reaction with 50% "catch" trials; the stimulus is known but there is 1 bit of uncertainty as to whether or not it will occur. In the absence of a better designation this will be referred to as an *A* reaction. The *B* reaction was used in Exp. III but unequal probabilities were introduced. In the two remaining experiments, pairs of equally probable but asymmetrical targets were used.

The *S*s wore headsets and listened to white masking noise of about 70 db. loudness. A momentary interruption of this noise served as the ready signal, with a stimulus light (when it occurred) appearing exactly 2 sec. later.

The time between the onset of the stimulus light and the breaking of contact between the stylus and the starting button (RT) was recorded, as well as the time between the initiation of the movement and initial contact of the stylus with a target or error plate (MT). The *S*s were told their average (RT + MT) time and number of errors after each block of trials. They were not aware of the analysis of response times into RT and MT.

RESULTS

Results for the Practice Session and for all experiments except Exp. III were highly consistent and will there-

TABLE 2

EXPERIMENTAL CONDITIONS FOR THE DIFFERENT EXPERIMENTS

	Target Pairs	Summary of Procedures
Practice Session	12 pairs of identical targets, on opposite sides of the starting button.	24 successive trials per target pair. Each target designated randomly and equally often on successive trials. Order of pairs randomized.
Exp. I	12 pairs of identical targets, on opposite sides of the starting button.	40 successive trials per target pair. One target designated preceding each trial, but the signal to respond given on only half the trials. Order of pairs randomized.
Exp. II	12 pairs of identical targets, on opposite sides of the starting button.	Same procedure followed as used for the Practice Session.
Exp. III	6 pairs of identical targets (C, D, H, I, J, K), on opposite sides of the starting button.	Same procedure followed as for the Practice Session and Exp. II except that one target of each pair designated on the average 50%, 75%, or 92% of the time over blocks of 24 trials.
Exp. IV	3 pairs of heterogeneous targets (C and I, D and J, H and K), each pair of equal difficulty but differing in *A* and *W*; both targets on the same side of the starting button.	24 successive trials per target pair. Tests made with both targets to the right and also with both to the left of the starting button.
Exp. V	3 pairs of heterogeneous targets (C and I, D and J, H and K), each pair of equal difficulty but differing in *A* and *W*; targets on opposite sides of the starting button.	24 successive trials per target pair. Tests made with each target of a pair alternately to the right and to the left of the starting button on successive blocks of trials.

fore be discussed together. The standard *B* reaction (Practice and Exp. II) will serve as the prototype.

Reaction times.—In all instances where pairs of identical targets were used RT data were very similar for the right and left targets. Accordingly, the data were combined for all identical pairs of targets.

The correlation between mean RT and ID value for the 12 movement conditions of Exp. II was found to be .79, which is statistically significant. RT increased slightly as movement amplitude (*A*) was increased, and as target width (*W*) was decreased, although only the former effects were significant ($p < .05$). However, the slope of the linear function relating the two variables was only 5.4 msec. per 1-bit change in ID (see Fig. 2). Results for Exp. I, IV, and V were also consistent in indicating that changes in ID values have a consistent but very small effect on RT.

Reaction times were slightly but significantly faster ($p < .01$) for Exp. I (*A* reaction with 50% stimulus uncertainty) than for Exp. II (two-alternative *B* reaction). The mean difference, over all 12 target conditions, was only 15 msec. Practice effects alone should have produced a difference in the opposite direction, however, so that the true RT difference between time and event uncertainty may be larger than 15 msec.

The only variable which had an appreciable effect on RT was the relative probability of the two alternative targets (degree of redundancy of a response sequence). These effects (from Exp. III) will be discussed later.

Movement times.—Changes in required movement characteristics had large as well as highly reliable effects on MT, in sharp contrast to their small effects on RT. The MT data from the Practice Session, and from Exp. I and II (see Fig. 2), permit a direct comparison with the earlier data from self-paced cyclical tasks (Fitts, 1954). It appears that the index of task difficulty (ID) provides equally as good a prediction of movement times for discrete responses as has been shown previously to hold for serial responses. As long as the level of stimulus uncertainty is held constant, the source of uncertainty (*A* reaction vs. *B* reaction) does not change the nature of the function. However, the slope of the function is less steep for discrete than for serial responses. Although individual differences were large, the effects of changes in target distances and widths were highly consistent within *S*s.

The movement times were of similar magnitude in Exp. I and II. Five of the six *S*s made faster movements under the *A* reaction condition while one *S* showed a very small (2 msec.) difference favoring the *B* reaction. Although the mean difference is statistically significant ($p < .05$), the effect is quite small relative to other experimental effects. Within each experiment about 99% of the variance in mean MT was accounted for by the ID ratio, the correlations between ID and MT for these two experiments being .995 and .994, respectively.

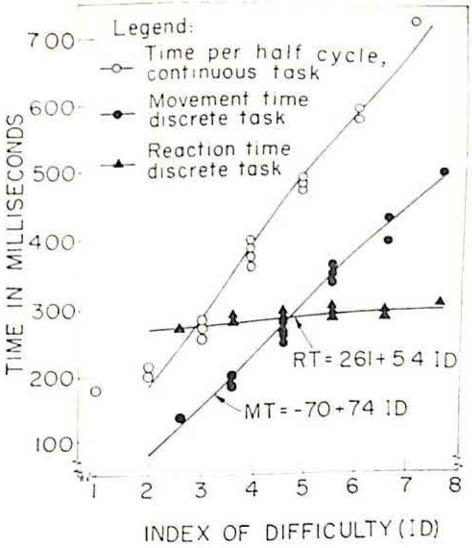


FIG. 2. Relation of reaction time and movement time to the index of task difficulty. (The solid data points show the relation of RT and MT data obtained in Exp. II, to ID—$\text{ID} = \log_2 2A/W$. The open circles show mean times per movement in performing a serial tapping task in which target amplitudes and widths were varied in the same manner as in the present experiments; see Fitts, 1954.)

Furthermore, the regression equations for predicting MT on the basis of a knowledge of required movement amplitude and accuracy are very similar for the two experiments as well as for the Practice Session. These three equations are:

$$\text{MT} = 74\ \text{ID} - 42\ \text{msec.} \qquad \text{(Practice Session)} \quad [6]$$

$$\text{MT} = 70\ \text{ID} - 63\ \text{msec.} \qquad \text{(Exp. I)} \quad [7]$$

$$\text{MT} = 74\ \text{ID} - 70\ \text{msec.} \qquad \text{(Exp. II)} \quad [8]$$

Examination of the constants of Equation 6 vs. 8, for comparable data taken at different stages of practice, suggests that practice resulted in a displacement of the function downward without a change in slope. Thus easy and difficult movements became

TABLE 3

MEAN REACTION AND MOVEMENT TIMES FOR COMPARABLE TARGET PAIRS IN EXP. II, IV, AND V

Target Pair	ID	Exp. II		Exp. IV		Exp. V	
		RT	MT	RT	MT	RT	MT
C and I	4.58	28	27	29	25	30	25
D and J	5.58	29	35	30	31	30	32
H and K	6.58	29	41	28	38	30	40

Note.—Data are for all *S*s and for right and left targets combined; times are shown to the nearest .01 sec.

slightly faster by similar absolute amounts.

The results of the analyses of movement durations for experiments in which pairs of dissimilar targets were used (Exp. IV and V) are summarized in Table 3. It can be seen that variations in target pairings had a very small effect on RTs. The actual times were very comparable to those found in Exp. II in which identical targets were placed at equal distances on opposite sides of the starting point. However, again it was found that MT increased progressively and by large amounts as a linear function of ID. Movement durations were similar in absolute magnitude to those found for homogeneous pairs of targets in Exp. II, the times for different movements varying over a range of about 150 msec.

MTs in Exp. IV and V were 33 and 23 msec. faster, respectively, than for comparable target conditions in Exp. II. Most or all of this increase could easily have been due to a continuing practice effect, however. Movements to the target on the side of the preferred hand were 20 and 11 msec. faster, respectively, than to the alternative side, whereas movements to the nearer and farther targets of a pair on the same side, equated for ID, differed by only 4 and 6 msec., respectively.

Errors.—Error data for the Practice Session and for Exp. I and II are summarized in Table 4. The *S*s failed to hit the target on approximately 10% of their responses, overshoot and undershoot errors being very nearly equal. Error frequencies were found to be essentially independent of movement amplitudes, but were more frequent for the two smaller than for the two larger targets. *A* responses, where *S* knew in advance which target he should try to hit, were no more accurate than *B* responses. (In fact, errors were slightly more frequent for the former.) The percentage of errors in Exp. IV and V were comparable, being 9.5% and 8.2%, respectively, and overshoot and undershoot errors were again almost exactly equal.

Redundant sequences.—The results from Exp. III, in which response series of three levels of redundancy were employed, differed from those obtained in other experiments in the series (see Table 5). As redundancy was increased as a consequence of one target of an identical pair being designated more frequently than the other, on the average, both RTs and MTs became slightly faster, for those responses that were made to the more probable target. Accuracy also became somewhat greater for the more probable target.

The differences between the RTs

TABLE 4

Percentage of Undershoot (U) and Overshoot (O) Errors for the Practice Session and for Exp. I and II

Target	Practice		Exp. I		Exp. II	
	U	O	U	O	U	O
A	4.2	0.7	1.7	4.2	5.6	5.6
B	3.5	2.1	2.5	5.8	2.8	5.6
C	6.2	4.9	10.0	6.7	5.6	3.5
D	4.2	4.9	7.5	7.5	10.4	6.9
E	4.9	3.5	2.5	4.2	2.8	2.1
F	2.8	3.5	7.5	8.3	3.5	4.2
G	7.6	5.6	4.2	8.3	2.8	4.2
H	8.3	6.9	7.5	10.8	9.0	10.4
I	1.4	1.4	2.5	4.2	5.6	2.1
J	4.2	5.6	8.3	5.8	2.8	4.2
K	6.2	4.9	8.3	10.8	2.8	9.7
L	6.2	5.6	8.3	9.2	6.9	7.6
M	5.0	4.1	5.9	7.2	5.0	5.5

Note.—The Practice Session and Exp. II involved 1,728 trials each; Exp. I involved 1,440 actual and 1,440 "false" trials. In Exp. I, three additional errors occurred as *Ss* lifted the stylus when a stimulus was not given.

to the alternative targets in each pair increased approximately linearly with redundancy, as has recently been reported by Fitts, Peterson, and Wolpe (1963) for two other tasks, and so did movement times. Although relatively small in magnitude, these effects were consistent for all *S*s and for all target pairs. Since time and errors are usually negatively correlated, the MT differences should have been even greater had *S*s achieved equal accuracy for the less frequent movements. Gains in speed and accuracy for the more frequent target of each pair, weighted for frequency, just about compensated for losses in speed and accuracy for the less frequent target.

DISCUSSION

Independence of reaction time and movement time.—The present results indicate that RT and MT can be influenced quite independently by (*a*) the degree of uncertainty regarding the stimulus to a

TABLE 5

Summary of Reaction Time, Movement Time, and Error Data for Movements in Two Directions in Exp. III

Percent Redundancy	Reaction Time (Msec.)				Movement Time (Msec.)				Percentage of Movements Missing Target			
	N	P	Diff.	Wtd. Aver.	N	P	Diff.	Wtd. Aver.	N	P	Diff.	Overall % Misses
0.0	293	279	14	286	342	339	3	340	10.0	8.3	1.7	9.1
18.9	315	282	33	290	354	341	13	344	12.0	9.7	2.3	10.3
58.7	326	268	58	271	366	343	23	344	22.2	8.2	14.0	9.4

Note.—P indicates a movement to the target on the side of the preferred hand and N a movement to the nonpreferred side; the target on the preferred side was made the most frequent target under conditions of 18.9% and 58.7% redundancy.

movement, and (*b*) the degree of uncertainty permitted in executing the movement. This finding supports the view that perceptual processes and motor response processes are relatively independent (see Broadbent, 1958; Welford, 1960), and that human information processing proceeds by a series of essentially independent steps.

Welford's (1958) conclusion that RT and not MT shows the major effect of aging offers strong parallel evidence for the separate nature of perceptual and motor information handling processes, as do results from related experiments of Leonard (1952), Singleton (1954), and Szafran (1951).

Apparently RT reflects the time required for perceptual or cognitive processes, and is determined in part by the preparations which *S* makes prior to a stimulus, such as those resulting from his knowledge of stimulus probabilities. Movement time, in contrast, appears to reflect the duration of motor system processes that are necessary for the control of the timing and patterning of a movement, and which begin after the decision is made to execute a movement. Advanced preparation for two-alternative movement processes does not appear to lengthen the time required to execute one of the movements, as compared to the situation where *S* is prepared in advance for only one response which has a 50% probability of being designated.

Several suggestions can be offered regarding processes which might utilize increased time to gain greater accuracy in the control of an output. As an illustration, if neural timing should involve feedback loops or reverberating circuits, then the relative effects of random biological noise in such loops should decrease as the timing interval increases.

Effects of redundancy and of time vs. event uncertainty.—The effects of differences in relative stimulus frequencies on RT, although small in absolute magnitude, are similar to those recently reported by Fitts, Peterson, and Wolpe (1963), and offer further support for a stimulus sampling theory of choice behavior. The close similarity of RT and error data for the *A* and *B* reactions, where 50% probability obtained in both instances, suggests that the effects of time uncertainty and of event uncertainty are closely equivalent.

The index of task difficulty.—Crossman (1960) and Welford (1960) have suggested that instead of actual target width, a corrected estimate of W, adjusted for errors, should be used in computing ID values. Applying such an empirical correction (so that $W = \pm 2SD$), the equation fitted to the data for Exp. I becomes

$$\mathrm{RT} = 74\ \mathrm{ID} - 57\ \mathrm{msec.} \qquad [9]$$

This indicates a slightly less efficient rate of processing information than does Equation 7. However, the correction has no appreciable effect on the correlation between ID and RT (changing it from .995 to .993). The main point for emphasis is an *S*'s remarkable ability for maintaining error rate relatively constant across a wide variety of task conditions, while varying the speed of his movements.

Welford has also suggested that an improved index of task difficulty might be

$$\mathrm{ID}' = \log 2\,\frac{A + .5W}{W} \qquad [10]$$

This suggestion is based in part on the observation that this definition reduces the numerical value of the first constant, a, in Equation 5, giving theoretical predictions of MT near zero for an ID′ of zero. Accordingly, Equation 5 was reevaluated for the data of Exp. I, using ID′ in place of ID. This expression was found to be

$$\mathrm{MT} = 72\ \mathrm{ID}' + 6\ \mathrm{msec.} \qquad [11]$$

Near zero MT is now predicted for values of $\mathrm{ID} = 0$ (in place of near zero MT for $\mathrm{ID} = 1$), and the correlation between these modified predictions and actual MTs was found to be .997. These results therefore support Welford's arguments.

It is obvious, however, that neither RT nor MT can ever approach zero in any real task. It also appears reasonable to expect that empirical determinations of human motor performance may indicate that capacity varies over the range of possible response amplitudes and tolerances. Since neither index has been derived formally from a theory, choice between them should rest on heuristic considerations. Some implications of the original index (Equation 3) are indicated below.

Human channel capacities.—Accepting Equation 3 as the definition of ID, and Equation 8 as providing a generalized empirical estimate of discrete movement times, the estimates of human performance capacity, as specified in Equation 4, would vary from about 22 bits per second for a value of ID = 2.5 (the least difficult movement studied), to just over 14 bits per second for ID = 7.5 (the most difficult movement studied). Thus it might appear that the motor system is relatively more efficient in producing low-information than high-information responses. Such estimates cannot be extended directly to self-paced cyclical motor tasks, however, since some additional uncertainty is introduced by the requirement that *Ss* process feedback data in repetitive tasks.

Earlier results (Fitts, 1954) indicate that actual performance rates in self-paced repetitive tasks vary from about 9.5 to 11.5 bits per second for ID values comparable to those studied here, and that maximum performance is attained for intermediate values of A, W, and ID. It is interesting to note that a linear combination of the equations for RT and MT would predict better performance for more difficult tasks as long as the effect of RT was weighted sufficiently to give a positive value for the intercept constant, a. Approximately a 4 to 1 weight for MT relative to RT would give a predicted zero intercept, or equivalent information rates at all difficulty levels.

The view that performance in self-paced cyclical tasks is a combined function of RT plus MT is, of course, an old one and, as mentioned earlier, agrees with the approach taken by Welford (1958) in analyzing age changes in perceptual-motor performance. It is significant that the average total time per response in the self-paced serial task (see Fig. 2) exceeded the present movement times by only about .1–.2 sec., and not by an amount as great as the present reaction times. This suggests that the processing of feedback data can to some extent overlap the processes involved in the control of subsequent movements, or else, as Woodworth (1899) and others have suggested, that a sequence of two or more movements may be programed in advance and executed independently of feedback. Such a finding is also consistent with Licklider's (1960) observation that *S*'s lag in a continuous tracking task is often less than reaction time data might suggest, and with the data from studies bearing on psychological refractory period delays (Welford, 1952) where the increase in RT to the second of two stimuli presented in rapid succession usually is less than a simple reaction time.

In summary, it appears that the processing of feedback data in serial or continuous tasks introduces some small delay relative to simple open-loop movements, but less delay than would be expected if every response involved a separate reaction time.

REFERENCES

Andriessen, J. J. Montageproeven bij diverse montagesnelheden (II). Report No. 13, 1960, Institute for Perception Research, Eindhoven, Netherlands.

Annett, J., Golby, C. W., & Kay, H. The measurement of elements in an assembly task: The information output of the human motor system. *Quart. J. exp. Psychol.*, 1958, **10**, 1–11.

Broadbent, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

Crossman, E. R. F. W. The information capacity of the human motor system in pursuit tracking. *Quart. J. exp. Psychol.*, 1960, **12**, 1–16.

FITTS, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *J. exp. Psychol.*, 1954, **47**, 381–391.

FITTS, P. M., PETERSON, J. R., & WOLPE, G. Cognitive aspects of information processing: II. Adjustments to stimulus redundancy. *J. exp. Psychol.*, 1963, **65**, 423–432.

LEONARD, J. A. Some experiments on the temporal relation between information and action. Unpublished doctoral dissertation, Cambridge University, 1952.

LICKLIDER, J. C. R. Quasi-linear operator models in the study of manual tracking. Part 3. In R. D. Luce (Ed.), *Developments in mathematical psychology.* Glencoe, Ill.: Free Press, 1960.

SHANNON, C. A mathematical theory of communication. *Bell Sys. tech. J.*, 1948, **27**, 379–423, 623–656.

SINGLETON, W. T. The change of movement timing with age. *Brit. J. Psychol.*, 1954, **45**, 166–172.

SZAFRAN, J. Changes with age and with exclusion of vision in performance at an aiming task. *Quart. J. exp. Psychol.*, 1951, **3**, 111–118.

VREDENBREGT, J. Metingen van tijden voor het monteren van pennen in gaten. Report No. 2, 1959, Institute for Perception Research, Eindhoven, Netherlands.

WELFORD, A. T. The psychological refractory period and the timing of high speed performance: A review and a theory. *Brit. J. Psychol.*, 1952, **43**, 2–19.

WELFORD, A. T. *Ageing and human skill.* Oxford: Oxford Univer. Press, 1958.

WELFORD, A. T. The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, 1960, **3**, 189–230.

WELFORD, A. T. Age changes in the times taken by choice, discrimination and the control of movement. *Gerontologia*, 1961, **5**, 129–145.

WOODWORTH, R. S. The accuracy of voluntary movement. *Psychol. Rev. monogr. Suppl.*, 1899, **3**(2, Whole No. 13).

(Received March 15, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 113–119

GSR CONDITIONING AND SENSITIZATION AS A FUNCTION OF INTERTRIAL INTERVAL[1]

WILLIAM F. PROKASY AND HARVEY C. EBEL

Pennsylvania State University

2 levels of intertrial interval (40 sec. vs. 20 sec.) were investigated in a GSR conditioning situation, with controls for sensitization incorporated. The dependent variables reported were: base, magnitude, latency, and recruitment. The primary results in acquisition were: (a) the conditioning operations led to greater magnitude and recruitment than did the sensitization operations; (b) greater magnitude and recruitment were associated with spaced, as opposed to massed, practice; and (c) sensitization and conditioning operations were differentiated across trials by a decrease in magnitude in the sensitization groups rather than by an increase in the conditioning groups. The primary results in extinction were: (a) the decrease in magnitude across extinction trials; (b) the longer latency of sensitized, as opposed to conditioned, responses; and (c) the greater magnitude associated with spaced as opposed to massed practice.

The results of several recent studies (Kimmel, 1959; Prokasy, Hall, & Fawcett, 1962; Stewart, Stern, Winokur, & Fredman, 1961) have drawn attention to the role of sensitization in classical GSR conditioning. Of particular concern are the conditions under which the effects specific to CS-UCS pairings can be separated from the influences of sensitization. Kimmel (1959) found that, with an intense auditory stimulus as CS, conditioning performance was no different from that obtained under sensitization procedures, i.e., performance to a tone when unpaired tone and shock are presented in an unpredictable order. In contast, with a low-intensity auditory CS, evidence of the effects of CS-UCS pairings was illustrated by the superior performance of the conditioning groups to its sensitization control. This superiority, however, was largely dissipated by the tenth paired trial. Stewart et al. (1961) employed a longer interstimulus interval (7.5 sec.) and measured anticipatory responses to shock stimulation. Though their results are subject to revision since they incorporated no controls for sensitization, they report an acquisition function which spontaneously dissipated by the sixteenth CS-UCS pairing. Prokasy, Hall, and Fawcett (1962), on the other hand, compared extinction performances of conditioning and sensitization groups and found no differences between them, i.e., no evidence of effects peculiar to CS-UCS pairings.

Although the cited studies vary widely in method, one major difference among them is the average duration of the intertrial interval (ITI). Kimmel (1959) and Stewart et al. (1961) employed an average ITI of over twice as long as that employed by Prokasy, Hall, and Fawcett (1962). It was the purpose of this study, therefore, to investigate the role of ITI in classical GSR conditioning,

---

[1] This study was supported by funds from a grant (G-18119) to the senior author from the National Science Foundation. The authors gratefully acknowledge the technical assistance of Barbara B. Prokasy.

This research was conducted while the junior author was a National Institute of Mental Health Predoctoral Research Fellow.

with controls for sensitization incorporated.

METHOD

Apparatus.—The apparatus for stimulus presentation and GSR measurement is described in detail elsewhere (Prokasy, Fawcett, & Hall, 1962). Briefly, the stimulus durations and temporal relationships were controlled by Hunter-Brown timers, the 1,000-cps tone CS was presented through earphones via a Grason-Stadler twin oscillator, the shock UCS was presented via an ac variac, and responses were amplified by a Fels dermohmmeter and recorded on an Esterline-Angus milliammeter operating at a paper speed of 24 in. per min.

Subjects.—A total of 120 volunteers, informed that shock would be employed, from introductory psychology classes at the Pennsylvania State University served as *Ss*. Of these, 8 were lost through adaptation to UCS, apparatus failure, or *E* error.

Design and procedure.—The basic design was a 2 × 2 factorial, with *Ss* assigned randomly to each of the four conditions. One independent variable was conditioning vs. sensitization and the other was massing (average ITI = 20 sec., with randomly ordered intervals of 10, 20, and 30 sec.) vs. spacing (average ITI = 40 sec., with randomly ordered intervals of 30, 40, and 50 sec.). Thus, the four experimental conditions were: conditioning-massed trials (C20), conditioning-spaced trials (C40), sensitization-massed trials (S20), and sensitization-spaced trials (S40). The two conditioning groups received 6 adaptation trials, 40 acquisition trials, and 10 extinction trials. The 40 acquisition trials were presented on a 75% irregular reinforcement schedule with an interstimulus interval of .5 sec. The two sensitization groups received 6 adaptation trials, 40 "acquisition" trials composed of an irregular order of 30 shock-alone trials and 10 tone-alone trials, and 10 "extinction" (tone-alone) trials. For all four groups, the tone-alone (test) trials during the acquisition phase occurred on Trial No. 2, 4, 12, 13, 15, 17, 22, 27, 33, and 35. Thus, the only procedural difference between the conditioning and sensitization groups was that, where the conditioning groups received a CS-UCS pairing, the sensitization groups received the UCS alone.

The *Ss* were seated in a partially sound-isolated cubicle 5 × 6 × 7 ft. high. Copper shock electrodes were placed on *S*'s left forefinger and GSR electrodes were placed on *S*'s right palm. The GSR electrodes were .75-in. zinc electrodes in plastic cups and established contact with *S* through a .125-in. layer of electrode paste. Shock level was determined for each *S* individually by starting at a low shock level and gradually turning it up until it was "highly annoying, but not painful." The resulting means and *SDs*, respectively, of voltage were: Group C20: 72.7, 18.2; Group C40: 70.8, 14.4; S20:67, 12.6; S40: 69.4, 13.4. Following this adjustment each *S* was instructed that he might receive either a tone or a shock stimulation, but that there was no necessary relationship between them. The *S* was told to maintain a passive attitude throughout the session, which could last as long as 40 min. (actually only about 20 min. for Groups C20 and S20).

After instructions were completed, *E* returned to the apparatus room, where he selected, by means of rotating a 12-segment cardboard disk, 1 of 12 teletype tapes which, when operated through a Western-Union teletype transmitter, programed both ITI and stimulus events. There were three such tapes for each of the four treatment combinations, the three tapes within a set differing only by virtue of random ITI orders. This procedure of assigning *Ss* to experimental conditions was employed in order to minimize *E* influence on *Ss* as a function of the treatment combinations. With the selection of the program tape, the session began.

The CS, a 70-db. (re .0002 dynes/cm^2) 1,000-cps tone, lasting 700 msec., was presented by means of a Grason-Stadler twin oscillator through Permoflux PDR-8 earphones, and was superimposed on a constant background of white noise generated through the earphones at 58 db. The UCS was a 200-msec. shock which began 500 msec. after CS onset in Groups C20 and C40.

Dependent variables.—Prokasy, Fawcett, and Hall (1962) employed measures of magnitude, amplitude, latency, and recruitment. The same measures were employed in this study, with the addition of base level. Magnitude is a measure of change in base which incorporates zero response (i.e., if there is no response, a score of zero is averaged with the changes recorded on other trials). Amplitude is a measure of change in base, given that a response has occurred: thus, responses of zero magnitude are not counted. Since analyses of amplitude yielded no substantial information beyond that obtained from the magnitude measure, the amplitude results will not be reported.

Recruitment is the amount of time elapsing between the onset of a response and its peak, while latency is measured from CS onset to response onset. Latency and recruitment were measured to the nearest 75 msec. Magnitude and amplitude were recorded in units of the square root of micromhos conductance change ($\sqrt{\Delta C}$). Measures of base, recorded in micromhos conductance, were taken during the first .9 sec. following CS onset on each trial. Changes in base occurring between .9 and 5.0 sec. following CS onset were counted as responses.

RESULTS

Since assignment of *S*s to treatments was done randomly following a reading of the instructions, resulting group *N*s differed. The *N*s for Groups C40, C20, S40, and S20, were, respectively, 24, 32, 26, and 30.

Although base and magnitude are easily measured, to obtain latency and recruitment requires that a response occur. To provide a common basis for all measures, analyses were conducted on those *S*s who responded a minimum of five times in the acquisition and extinction phases (i.e., on 50% of the test trials within each phase). To this end, test trials were pooled into units of trial pairs. If *S* made a response on only one trial of the pair, latency and recruitment values of that trial were treated as averages for the trial pair. If he made a response on both trials of the pair, latency and recuitment values were averaged for the trial pair. In contrast, base and magnitude values were averaged over both trials, regardless of whether there were one or two responses. As a result of this procedure, the *N*s employed in the analyses varied depending upon the phase of training. In addition, to simplify the analyses *S*s were randomly sampled out where necessary to provide cell proportionality. The *N*s employed are reported with each analysis.

Though the basic analyses were performed on *S*s meeting the frequency criterion, it is apparent that, with respect of the magnitude and base indexes, such a criterion imposes *S* selection; i.e., *S*s responding on less than half of the test trials were not retained for analyses. In order to determine whether this form of selection biased the results, analyses of magnitude and base were made employing the original *N*s reported above, with the exception that the data of two randomly selected *S*s, each in Groups C20 and S40, were deleted in order to achieve cell proportionality. Unless otherwise reported, the results of these analyses were the same as those reported below.

Acquisition.—After determining which *S*s met the frequency criterion and then restoring cell proportionality, the available *N*s for analysis of acquisition data in Groups C40, C20, S40, and S20 were, respectively, 24, 25, 24, and 25.

Table 1 presents means for the four GSR attributes across the five pairs of acquisition trials, while Table 2 summarizes the corresponding analyses of variance. Neither of the independent variables substantially affected base conductance, but there was a significant overall decrease in conductance across the five pairs of test trials. The introduction, following adaptation, of two conditioning trials in Groups C40 and C20 and two shock trials in Groups S40 and S20 resulted in an increase in conductance for each *S* from the last pair of adaptation test trials to the first pair of acquisition test trials, the overall mean changing from 18.9 to 21.9 micromhos.

The conditioning groups yielded higher $\sqrt{\Delta C}$ than did the sensitization groups. The significant Trials effect reflects an overall *decrease* in magnitude across the five pairs of test trials, but the significant Trials × Con.-Sen.

TABLE 1

MEAN BASE, MAGNITUDE, LATENCY, AND RECRUITMENT OVER FIVE BLOCKS OF TWO ACQUISITION TRIALS EACH

Index	ITI	*N*	Conditioning Trial Blocks					Sensitization Trial Blocks				
			1	2	3	4	5	1	2	3	4	5
Base (conduct.)	Spaced	24	19.89	19.64	19.62	18.87	18.65	25.61	25.01	24.20	23.53	23.38
	Massed	25	20.65	21.47	21.00	20.87	20.78	21.69	21.30	21.56	20.92	20.57
Magnitude ($\sqrt{\Delta C}$)	Spaced	24	1.001	1.048	1.051	1.065	1.017	.883	.757	.669	.646	.512
	Massed	25	.804	.876	.869	.850	.796	.837	.670	.726	.652	.454
Latency (sec.)	Spaced	24	2.053	2.026	2.094	2.098	2.075	1.982	2.123	2.117	2.122	2.292
	Massed	25	2.334	2.127	2.219	2.166	2.165	2.091	2.261	2.184	2.232	2.163
Recruitment (sec.)	Spaced	24	2.544	2.616	2.491	2.593	2.513	2.306	2.189	2.240	2.306	2.072
	Massed	25	2.174	2.340	2.105	2.282	2.190	2.285	2.148	2.082	1.995	1.872

Note.—Within-cell *SD*s varied from 8.83 to 20.44 for base, from .299 to .574 for magnitude, from .222 to .498 for latency, and from .342 to .796 for recruitment.

interaction indicates that the decrease is quite large in the sensitization groups and negligible, if at all, in the conditioning groups. Response magnitude was at its peak on the first pair of test trials; thus, there was no indication of a gradual rise (or an "acquisition" function) peculiar to the conditioning groups.

The nonsignificant ITI effect on magnitude requires further comment. Incorporating data from all *S*s (including those responding on fewer than half of the test trials), mean magnitude for the spaced and massed conditions was, respectively, .84 and .69 $\sqrt{\Delta C}$. These means are significantly different, $F(1, 104) = 4.45$, Error $MS = .688$. Since the difference in the two analyses rests in the inclusion, in this latter analysis, of five *S*s in Group S20 who failed to respond at least five times in the 10 test trials (plus six *S*s *randomly* selected out of Group C20 to restore all proportionality) the nonsignificant

TABLE 2

ANALYSIS OF VARIANCE OF THE ACQUISITION PHASE

Source	*df*	Base		Magnitude		Latency		Recruitment	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between *S*s	97								
Con.-Sen. (C-S)	1	818.38		7.988	12.33***	.050		6.719	9.11***
ITI	1	71.16		1.525	2.35	1.057	3.03	7.298	9.90***
C-S × ITI	1	693.163		.905	1.38	.166		1.091	1.47
Error (b)	94	973.943	107.32***	.648	8.93***	.345	5.21***	.737	3.78***
Within *S*s	392								
Trials (T)	4	23.45	2.58*	.478	6.85***	.046		.497	2.54*
T × C-S	4	4.76		.468	6.44***	.305	4.57***	.426	2.18
T × ITI	4	6.77		.023		.146	2.19	.094	
T × C-S × ITI	4	1.85		.023		.088	1.31	.106	
Error (w)	376	9.075		.073		.067		.195	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

distributional effect on magnitude appears to have resulted from a forced increase in mean response level for Group S20. As a result, a separate comparison of Groups C40 and C20 was made, and the spaced ITI yielded a $\sqrt{\Delta C}$ significantly larger than did the massed ITI, $F\ (1, 52) = 4.15$.

There were no significant main effects with latency as the dependent variable, although latency was somewhat greater under the massed condition. The significant Trials × Con.-Sen. interaction reflects a general latency decrease in the conditioning groups and a latency increase in the sensitization groups, the net effect of which is to nullify any overall Trials effect.

The conditioning and sensitization groups were differentiated with the recruitment index, the conditioning operations resulting in greater recruitment time (i.e., longer time for the measured response to reach its peak). The ITI variable was just as effective, with spacing (40-sec. ITI) resulting in significantly longer recruitment time than massing (20-sec. ITI). Recruitment decreased significantly across trials, with the greatest decrements observed in the sensitization groups, although the Trials × Con.-Sen. interaction was not statistically significant.

Extinction.—The available *N*s for the extinction data were 22, 22, 15, and 14 for, respectively, Groups C40, C20, S40, and S20. Table 3 provides the means of all indexes across the five pairs of extinction trials, while Table 4 summarizes the corresponding analyses of variance. It should be noted that 9 *S*s in Group S40 and 11 in Group S20 failed to meet the frequency criterion while only 2 in Group C40 and 5 in Group C20 did not meet the frequency criterion. This attrition of 7 *S*s in the conditioning groups and 20 in the sensitization groups in itself indicates a continued differentiation of sensitization and conditioning groups. The difference in proportion of *S*s not meeting the frequency criterion (7/54 vs. 20/54) is significant beyond the .01 level ($\chi^2 = 8.34$).

Base conductance continued to decrease throughout extinction trials, as indicated by the values in Table 3 and the significant Trials effect shown in Table 4. While the Con.-Sen. and ITI variables have no significant overall main effect, these two variables do interact, since base con-

TABLE 3

MEAN BASE, MAGNITUDE, LATENCY, AND RECRUITMENT OVER FIVE BLOCKS OF TWO EXTINCTION TRIALS EACH

Index	ITI	Conditioning ($N = 22$) Trial Blocks					Sensitization ($N = 15$) Trial Blocks				
		1	2	3	4	5	1	2	3	4	5
Base (conduct.)	Spaced Massed	19.10 22.52	18.32 22.02	17.97 21.73	17.23 21.69	16.71 21.46	28.40 19.72	28.40 19.89	28.97 17.95	28.78 17.52	27.09 17.55
Magnitude ($\sqrt{\Delta C}$)	Spaced Massed	1.046 .869	.966 .750	1.012 .784	.885 .716	.899 .586	.605 .561	.551 .537	.567 .403	.526 .366	.496 .344
Latency (sec.)	Spaced Massed	2.076 2.117	2.117 2.206	2.046 2.158	2.039 2.085	2.122 2.063	2.198 2.195	2.208 2.308	2.360 2.383	2.358 2.363	2.513 2.418
Recruitment (sec.)	Spaced Massed	2.582 2.192	2.422 2.071	2.485 2.098	2.538 2.284	2.472 2.103	2.018 2.125	2.128 1.903	2.476 2.065	2.495 2.008	2.396 2.080

Note.—Within-cell *SD*s varied from 8.21 to 18.63 for base, from .291 to .566 for magnitude, from .217 to .456 for latency, and from .440 to 1.104 for recruitment.

TABLE 4

Analysis of Variance of the Extinction Phase

Source	*df*	Base		Magnitude		Latency		Recruitment	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between *S*s	73								
C-S	1	1126.21	1.26	11.330	20.05***	4.625	13.85***	2.165	1.94
ITI	1	232.50		2.828	5.00**	.081		9.293	8.34***
C-S × ITI	1	4259.96	4.78**	.288		.036		.157	
Error (b)	70	890.09	276.85***	.565	7.79***	.334	2.08***	1.113	4.55***
Within *S*s	296								
T	4	34.26	10.65***	.418	5.75***	.114		.363	1.48
T × C-S	4	1.75		.028		.275	1.71	.274	1.11
T × ITI	4	4.33	1.34	.048		.080		.156	
T × C-S × ITI	4	11.47	3.56*	.028		.006		.321	1.31
Error (w)	280	3.22		.073		.160		.244	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

ductance is higher for the spaced, as opposed to the massed, sensitization group while it is higher for the massed, as opposed to the spaced, conditioning group. When incorporating all *S*s, the same trend exists, but it is not significant, $F\ (1, 104) = 2.04$. The difference in these two test results because the attrition of 11 *S*s in Group S20 resulted in a marked increase in base conductance; i.e., the 11 *S*s had lower base conductances than did the remaining 19 *S*s. The significant Trials × Con.-Sen. × ITI interaction, also significant in the analysis with all *S*s, is difficult to locate with any assurance in the means of Table 3.

The conditioning operations continued to result in greater magnitude than did the sensitization operations, and, as would be expected, magnitude decreased across extinction trials.

Response latency varied in extinction as a function of the conditioning-sensitization variable, with conditioning resulting in a shorter response latency. This result, significant at well beyond the .001 level (Table 4), appears to stem primarily from the latency increments across extinction trials in the sensitization groups, with the conditioning groups maintaining a relatively constant latency. This differential result was not, however, sufficiently large to produce a significant Trials × Con.-Sen. interaction.

Although the difference in recruitment as a function of the conditioning-sensitization variable obtained in acquisition dissipated in extinction (largely because of the sizable increment in Group S40, as contrasted with slight decrements for the other three groups), the long ITI continued to yield a significantly greater recruitment than did the short ITI.

Discussion

The major purposes of this experiment were (*a*) to investigate the ITI variable in GSR conditioning and (*b*) to obtain further information on factors which differentiate conditioning and sensitization performance. Our results indicate, first, that massing reduces magnitude and recruitment, consistent with findings on other classically conditioned response systems, such as the eyelid response (e.g., Prokasy, Grant, & Myers, 1958; Spence & Norris, 1950).

Second, not only was Group C40 superior to Group C20, Group C20 was differentiable from the two sensitization control groups. Thus, while the degree to which sensitization and conditioning operations result in differential performance is positively related to ITI duration, it is possible to obtain GSR conditioning with ITIs as short as 20 sec. Since recovery from the UCR can require as long as, or longer than, 20 sec., this particular value may represent a practical lower ITI bound in GSR conditioning if reasonably "clean" measures are to be obtained on successive trials. Our results suggest, furthermore, that with short ITIs, sensitization and conditioning performances were not distinct during early acquisition trials, but emerge as distinct mainly because response magnitude *decreased* across trials in the sensitization groups. There was, therefore, no evidence of the acquisition function (i.e., gradually increasing strength as a function of specific CS-UCS pairings) typically found with other classically conditioned response systems (e.g., Prokasy et al., 1958; Spence & Norris, 1950).

Performance in the conditioning groups remained constant across the 40 acquisition trials. This contrasts with the gradually *decreasing* magnitude across even fewer trials reported elsewhere (Kimmel, 1959; Stewart et al., 1961). Three variables which, singly or in combination, may account for the discrepancy are: the intermittent reinforcement schedule and short interstimulus interval employed in the present study and any difference in subjective UCS intensity.

Although it was a highly reliable measure, base conductance was not greatly affected by the sensitization vs. conditioning or the massing vs. spacing manipulations. A general trend, however, was a decrease in conductance across acquisition and extinction.

Since neither latency nor recruitment are generally measured in GSR conditioning, there is little in the literature as a basis for comparison. It is apparent that these measures are sensitive to independent variable manipulation (see also Prokasy, Fawcett, & Hall, 1962), and that they do not necessarily covary with independent variable effects on magnitude. For example, Prokasy, Fawcett, and Hall (1962) found that both recruitment and latency in extinction were significantly greater with long interstimulus intervals than with short ones, though there were no magnitude or amplitude differences. There is a discrepancy between the present study and that of Prokasy, Fawcett, and Hall (1962) in that latency increased across extinction trials in the earlier study, but did not in the present study. This discrepancy is best explained by the fact that the latency changes observed by Prokasy, Fawcett, and Hall were associated primarily with interstimulus intervals longer than 500 msec., whereas that was the interstimulus interval of the present study. In any event, that latency and recruitment do vary with independent variable manipulation is sufficient rationale to include these measures as a further step toward our understanding of GSR conditioning.

REFERENCES

KIMMEL, H. D. Amount of conditioning and intensity of the conditioned stimulus. *J. exp. Psychol.*, 1959, **58,** 283–288.

PROKASY, W. F., FAWCETT, J. T., & HALL, J. F. Recruitment, latency, magnitude and amplitude of the GSR as a function of interstimulus interval. *J. exp. Psychol.*, 1962, **64,** 513–518.

PROKASY, W. F., GRANT, D. A., & MYERS, N. A. Eyelid conditioning as a function of unconditioned stimulus intensity and intertrial interval. *J. exp. Psychol.*, 1958, **55,** 242–246.

PROKASY, W. F., HALL, J. F., & FAWCETT, J. T. Adaptation, sensitization, forward and backward conditioning, and pseudo-conditioning of the GSR. *Psychol. Rep.*, 1962, **10,** 103–106.

SPENCE, K. W., & NORRIS, E. B. Eyelid conditioning as a function of intertrial interval. *J. exp. Psychol.*, 1950, **40,** 305–313.

STEWART, M. A., STERN, J. A., WINOKUR, G., & FREDMAN, S. An analysis of GSR conditioning. *Psychol. Rev.*, 1961, **68,** 60–67.

(Received February 11, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 120–123

INFLUENCE OF THE SPATIAL RELATIONSHIPS BETWEEN THE CUE, REWARD, AND RESPONSE IN DISCRIMINATION LEARNING[1]

ROBERT E. MILLER AND JOHN V. MURPHY[2]

School of Medicine, University of Pittsburgh

This experiment, using 17 monkeys, investigated the effect on discrimination learning of reward separation from cue and response, and response separation from cue and reward. The results indicated that a spatial displacement of the reward from the cue and response did not seriously impair learning as compared to spatial contiguity between cue, response, and reward. Response separation, however, significantly impaired learning. It was concluded that efficiency of discrimination learning is markedly affected by spatial arrangements of cue, response, and reward.

The spatial arrangements of the cue, response, and reward have been identified in a series of recent experiments as a very potent variable in the learning of a discrimination (McClearn & Harlow, 1954; Meyer, Polidora, & McConnell, 1961; Miller & Murphy, 1956; Murphy & Miller, 1955, 1958; Schuck, Polidora, McConnell, & Meyer, 1961). The evidence suggests that spatial factors may be one of the most critical of variables though it has been largely ignored in most of the traditional experiments, theories, and apparatus of the experimental psychologist. Even slight deviations from optimal spatial relationships in a learning situation impair or, indeed, preclude discrimination learning.

While the published reports have unanimously agreed that spatial displacement of the discriminanda from the manipulanda and reward impedes learning, other conditions of spatial discontiguity did not appear to retard learning of a simple discrimination (Murphy & Miller, 1958). The present experiment was designed to examine more closely the effects on discrimination learning of separating the reward from cue and response (Cond. A) and separating the response from cue and reward (Cond. B) in comparison with the condition of spatial contiguity between cue, response, and reward (Cond. C). A fourth condition, (D) center feeding, was also tested. In this condition the cue and response were contiguous but reward for a correct response was delivered in a central location regardless of the position, right or left, of the positive stimulus and response. These four conditions of spatial arrangements are shown in Fig. 1.

METHOD

Subjects.—Seventeen experimentally naive male rhesus monkeys were used in this study. Four of these animals ranged in weight from 18 to 33 lb. The remaining 13 monkeys weighed between 10 and 14 lb. The animals were assigned randomly to groups prior to the experiment with the restriction that one of the larger monkeys be placed in each of the four groups.

Apparatus.—The apparatus has previously

---

[1] This experiment was supported in part by a research grant (M-487) from the National Institute of Mental Health, National Institutes of Health, United States Public Health Service, and in part by the Commonwealth of Pennsylvania.

[2] Now associated with the Thiokol Chemical Corporation, Humetrics Division, Los Angeles, California.

been described in detail (Murphy & Miller, 1958). The Wisconsin General Test Apparatus (Harlow & Bromer, 1938) was the basic unit of equipment. A special two-shelf formboard was constructed with the upper shelf located 7 in. above the bottom shelf. Two interchangeable trays were designed to fit the shelves and each was equipped with a pair of magnetic cabinet latches located 14 in. apart. One of the trays had a foodwell placed 2 in. directly in front of each of the magnetic latches. The second tray had only a single foodwell placed midway between the magnetic latches. The foodwells were covered by Micarta slides which could be retracted by *E* to expose the reward in the well when a correct choice occurred.

The response blocks were 2 × 2 × $\frac{3}{4}$ in. varnished hardwood blocks which were placed on the tray and could be moved freely by the monkey. The discriminative stimuli were 2 × 2 × $\frac{3}{4}$ in. blocks attached to a steel corner brace. These stimulus objects were held in position on the tray by placing the brace on the appropriate magnetic cabinet latch. When they were in position the stimuli were held $\frac{3}{4}$ in. above the surface of the tray and 1$\frac{1}{2}$ in. in front of the magnets. The faces of the discriminative stimuli were divided by a diagonal into equally sized black and white triangles. The positive stimulus for all of the animals was the stimulus block with the base of the white triangle at the bottom. The correct response which was accompanied by a food reward was displacement of the response block located on the same side of the apparatus as the positive stimulus.

Procedure.—In the preliminary training phase the animals were first taught to displace a response block to obtain a raisin located directly beneath it. Each animal was then trained under the condition of response-reward separation indicated by the group to which it had been randomly assigned. This preliminary training was continued until the animal was responding and taking the reward with no hesitation. No stimulus blocks of any kind were presented during this adaptation period.

In the discrimination learning situation, an opaque screen was lowered, stimulus objects and response blocks were positioned for the succeeding trial, and the appropriate foodwell was baited with a raisin or piece of grape. The opaque screen was then raised and the formboard was pushed within reach of the monkey. If the animal displaced the response block on the same side as the positive stimulus, the appropriate foodwell cover was immediately withdrawn by *E*. As noted in previous reports, this movement caused no apparent concern to any of the animals even in the preliminary training phase. In addition, this procedure occasioned no delay of reward since the cover was withdrawn as the response was being completed.

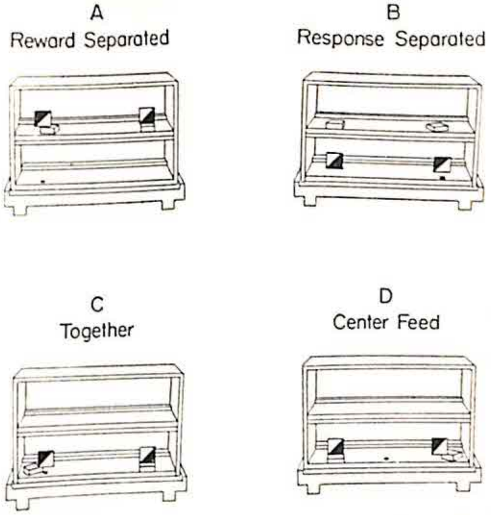


FIG. 1. The four conditions of spatial contiguity presented during this experiment. (In each case the drawing represents the appearance of the apparatus following a correct response. The dark spot in each panel is the open and baited foodwell. Note that the appropriate response block has been displaced from its original position, i.e., squared with the front of the tray.)

The conditions of cue, reward, and response relationships for the four groups were as follows: (Cond. A) reward-separated condition, in this case the cue and response were on the top shelf and the reward for a correct choice was presented 7 in. directly below (Fig. 1A); (Cond. B) response-separated condition, the cue and reward were contiguous on the bottom shelf while the response blocks were located on the top shelf (Fig. 1B); (Cond. C) nonseparated condition, in this arrangement, the stimulus blocks, response blocks, and reward were all contiguously located on the lower shelf as in the usual discrimination problem (Fig. 1C); (Cond. D) center-fed condition, cue and response contiguous on the bottom shelf with the reward located in the foodwell midway between the stimuli (Fig. 1D).

Each animal was given 50 trials daily; 25 trials in a morning test session and 25 in an afternoon session. Tests were conducted daily except Sunday throughout the investigation. Position of the positive object for each trial was systematically varied according to the Gellerman series (Gellerman, 1933).

Group C was terminated after 1,200 trials since all of the animals were performing above 90% correct responses. The center-feed group (D) was terminated after 1,500 trials when three of the four animals were performing above 90% correct responses and the fourth monkey had maintained a position habit for more than 1,000 trials and was, consequently, performing consistently at the 50% level. The remaining groups (A and B) were tested for a total of 2,700 trials.

RESULTS AND DISCUSSION

The learning curves for the four groups of animals are shown in Fig. 2. The data for the first 1,200 trials were analyzed by Bartlett's test (Bartlett, 1937) for homogeneity of variance. The test revealed that the group variances were not homogeneous ($p < .01$). Therefore, the error scores were evaluated by distribution-free methods.

The Kruskal-Wallis *H* test (Kruskal & Wallis, 1952) indicated that the four groups differed significantly in terms of errors in 1,200 learning trials ($p < .01$). The comparisons between specific conditions were made by Mann-Whitney tests (Mann & Whitney, 1947). The results of these comparisons are shown in Table 1.

The several recent experiments cited in the introduction have unequivocally demonstrated that spatial separation of the discriminanda from both the response and reward by even a very small distance severely retards learning of very simple discriminations. However, contiguity between the cue and either response or reward permits simple discrimination learning to occur (Murphy & Miller, 1958).

The present study was designed to differentiate between the cue-reward contiguity and cue-response contiguity conditions in a more complex discrimination problem. The results clearly indicate that spatial separation of the response from cue and reward retarded learning in the monkey. In the response separated group three of the four animals set up rigid position habits within 400–500 trials and did not deviate from this pattern for the remaining 2,200 trials.

The remaining three groups of animals, together, center-feed, and reward separated, all demonstrated considerable improvement in performance over the initial 1,200 trials. Learning under the

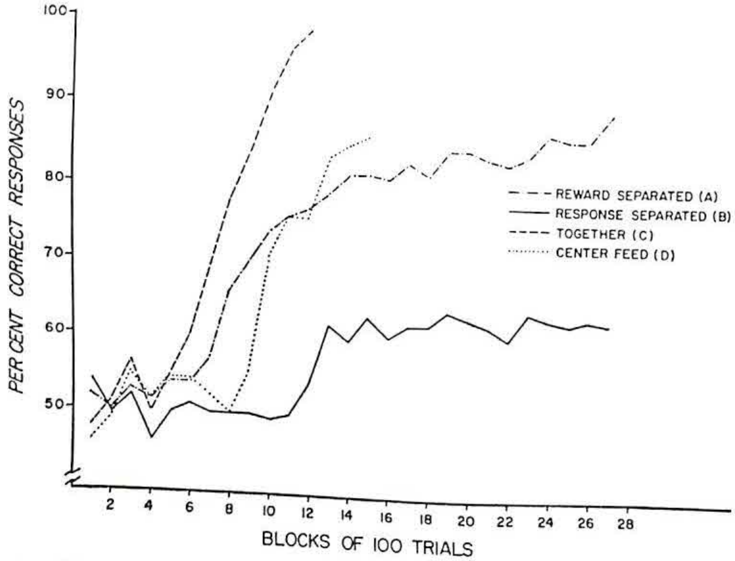


FIG. 2. The discrimination performance of the four groups of monkeys.

TABLE 1

INTERGROUP COMPARISONS FOR THE FIRST 1,200 TRIALS

Group	*p* (Two-Tailed)
Together vs. Center	<.04
Together vs. Response Separated	<.04
Together vs. Reward Separated	*ns*
Center vs. Response Separated	<.04
Center vs. Reward Separated	*ns*
Response Separated vs. Reward Separated	<.02

Note.—These analyses were made by the Mann-Whitney test.

together condition was significantly superior to the center-feed condition in terms of errors in 1,200 trials. Since these two test situations were identical in every respect other than site of the reward, the difference in rate of learning might be attributable to the ambiguity in the center-feed condition of receiving the reward in the same locus regardless of the side on which the correct response was performed.

The results of this experiment emphasize the importance of spatial factors in discrimination learning. Earlier studies have shown that even slight displacement of the stimulus from the location of the rewarded response drastically retards or precludes very simple discrimination learning. The present experiment has demonstrated that in more complex discriminations it is essential that there be spatial contiguity between the stimulus and the response.

REFERENCES

BARTLETT, M. S. Some examples of statistical methods of research in agriculture and applied biology. *J. Roy Statist. Soc. Suppl.*, 1937, **4,** 137–170.

GELLERMAN, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.*, 1933, **42,** 206–208.

HARLOW, H. F., & BROMER, J. A. A test apparatus for monkeys. *Psychol. Rec.*, 1938, **2,** 434–436.

KRUSKAL, W. H., & WALLIS, W. A. Use of ranks in one-criterion variance analysis. *J. Amer. Statist. Ass.*, 1952, **47,** 583–621.

MCCLEARN, G. E., & HARLOW, H. F. The effect of spatial contiguity on discrimination learning by rhesus monkeys. *J. comp. physiol. Psychol.*, 1954, **47,** 391–394.

MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18,** 50–60.

MEYER, D. R., POLIDORA, V. J., & MCCONNELL, D. G. Effects of spatial S-R contiguity and response delay upon discriminative performances by monkeys. *J. comp. physiol. Psychol.*, 1961, **54,** 175–177.

MILLER, R. E., & MURPHY, J. V. Discrimination learning with vertical vs. horizontal stimulus relationships. *J. comp. physiol. Psychol.*, 1956, **49,** 80–83.

MURPHY, J. V., & MILLER, R. E. The effect of spatial contiguity of cue and reward in the object-quality learning of rhesus monkeys. *J. comp. physiol. Psychol.*, 1955, **48,** 221–224.

MURPHY, J. V., & MILLER, R. E. Effect of the spatial relationship between cue, reward, and response in simple discrimination learning. *J. exp. Psychol.*, 1958, **56,** 26–30.

SCHUCK, J. R., POLIDORA, V. J., MCCONNELL, D. G., & MEYER, D. R. Response location as a factor in primate pattern discrimination. *J. comp. physiol. Psychol.*, 1961, **54,** 543–545.

(Received February 25, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 124–126

INTERACTION OF AROUSAL AND RECALL INTERVAL IN NONSENSE SYLLABLE PAIRED-ASSOCIATE LEARNING[1]

LEWIS J. KLEINSMITH AND STEPHEN KAPLAN

University of Michigan

Using meaningful paired associates as stimuli it has been shown previously that due to the phenomenon of perseverative consolidation high arousal associates show stronger permanent memory and weaker immediate memory than low arousal associates. The present experiment was designed to show that this phenomenon is independent of the association values or unique qualities of the words involved. While recording skin resistance as a measure of arousal, 36 Ss were presented 6 (0% association value) nonsense syllables paired with single digit numbers. Ss were tested at 2 min., 20 min., or 1 wk. The results confirmed the hypothesis ($p = .01$). Nonsense syllable paired associates learned under low arousal exhibited high immediate recall and rapid forgetting. High arousal associates exhibited a marked reminiscence effect, low immediate recall and high permanent memory.

Basing their argument on the hypothesis of reverberating neural circuits, Kleinsmith and Kaplan (1963) predicted that arousal and recall interval would interact in their influence on paired-associate learning. This prediction was confirmed; high arousal learning showed a marked reminiscence effect, that is, poor immediate recall but high permanent memory, while low arousal learning showed a typical forgetting curve pattern. In that study words of varying arousal qualities were used as stimuli.

The present study was designed to test the independence of the phenomenon from the association values or other qualities unique to the individual stimuli. Nonsense syllables of zero association value (Hilgard, 1951) were selected as stimuli in order to obtain random arousal effects. In this way both the high and low arousal curves would contain essentially the same items and thus any differential behavior in learning could not be explained in terms of qualities unique to the stimuli.

METHOD

Subjects.—The *S*s were 36 University of Michigan undergraduates obtained from introductory courses in psychology. They were run in three subgroups of 12 *S*s each.

Procedure.—The *S*s were given a single learning trial with a list of six nonsense syllable-number pairs. The following six 0% association value nonsense syllables were used: CEF, QAP, TOV, JEX, LAJ, DAX. The response items were single digits from 2 to 7.

A slide projector with an exposure time of 4 sec. was used to present the stimuli. During the training trial *S* first saw the nonsense syllable alone, and then saw the nonsense syllable repeated with a single digit response. To separate the arousal effects from one set of pairing to the next, two slides containing four colors each were inserted between the paired associates and *S* was instructed to name the colors. (Red, green, orange, blue, black, and yellow were used randomly on these slides.) The *S* was instructed to "concentrate carefully on both colors and nonsense syllable-number pairs" as he called them out loud, but to avoid rehearsal *S* was not specifically told that he would be tested for recall.

---

[1] This investigation was supported in part by Public Health Service Research Grant M4239 from the National Institutes of Health, Public Health Service, to E. L. Walker. The technical assistance of Dale E. Guckenberger is gratefully acknowledged.

In order to determine the specific arousal effects of each nonsense syllable, skin resistance was recorded during learning. The electrodes were of the zinc variety described by Lykken (1959). To insure constancy of conditions, electrodes and recording apparatus were also used during the recall sessions.

During the recall session *S* was instructed to indicate the correct number for each nonsense syllable as it appeared, and to guess if he was uncertain. The correct numbers were not repeated. Colors were used as an interpolated task as before.

Design.—The three groups were tested one at each of three recall intervals: immediate (about 2 min.), 20 min., and 1 wk.

To correct for serial order effects six different training lists were used, each list given to two *S*s in each group. The lists were designed so that each of the six nonsense syllables appeared once in each position in the list (Fisher & Yates, 1938). The order of the recall lists was varied in the same manner.

Data analysis.—Any drop in *S*'s skin resistance which occurred within 4 sec. of presentation of a nonsense syllable was considered an arousal deflection; these were converted to percent deflections. Each *S*'s six GSR deflections were then ranked. (In case of ties, a deflection occurring at a low level of absolute skin resistance was considered higher arousal than a similar deflection occurring at a higher absolute level.) The three highest deflections for each *S* were designated "high arousal learning," and the three lowest were designated "low arousal learning."

RESULTS

Table 1 shows the frequency of the six nonsense syllables for both high and low arousal at the three times tested. No systematic trends are present in the distribution of items which could account for the differences in behavior of high and low arousal learning.

Figure 1 illustrates the relationship between high and low arousal recall as a function of time. At immediate recall, numbers associated with low arousal nonsense syllables are recalled four times as often as numbers associated with high arousal nonsense syllables. The capacity to recall numbers associated with low arousal nonsense syllables decreases as a function of time in a characteristic forgetting curve pattern. On the other hand, the capacity to recall numbers associated with high arousal nonsense syllables shows a considerable reminiscence effect. After 20 min., the increase is 100%, and by a week it has increased 200%.

TABLE 1

NUMBER OF TIMES EACH NONSENSE SYLLABLE PRODUCED A HIGH OR LOW AROUSAL REACTION

	High Arousal			Low Arousal		
	2 Min.	20 Min.	1 Wk.	2 Min.	20 Min.	1 Wk.
TOV	5	5	6	7	7	6
CEF	7	6	8	5	6	4
QAP	8	7	6	4	5	6
DAX	6	5	6	6	7	6
LAJ	5	8	4	7	4	8
JEX	5	5	6	7	7	6

Analysis of variance (Lindquist, 1953) for correct responses confirms the trends in the figure at $p < .01$, $F(2, 33) = 5.74$. This significant interaction is primarily attributable to the effect of the immediate condition, $F(1, 11) = 10.81$, $p < .01$, and the

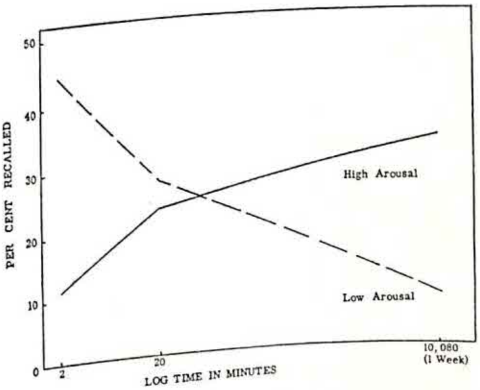


FIG. 1. Differential recall of nonsense syllable paired associates as a function of arousal level.

1-wk. condition, $F\ (1,\ 11) = 7.24$, $p < .025$. Thus at immediate recall, low arousal learning is significantly greater than high arousal learning; after 1 wk. the situation has reversed and high arousal recall is significantly better than low arousal recall.

Since there is some tendency for adaptation of GSRs during presentation of a list, the possibility arises that serial order interacts with the arousal classification. Thus it would be expected that the first two items would tend to fall in the high arousal category, while the last two items would be in the low arousal category. This is not borne out by the data. Analysis of recall as a function of interpolated interval and arousal classification, plotted separately for pairs of consecutive items, is essentially identical to the results shown in Fig. 1. Furthermore, if the first item is omitted from the analysis (since it contributes largely to the high arousal category) the results still remain like those in Fig. 1.

Discussion

Since the high and low arousal curves contain essentially the same items, the results show that differential recall ability can be predicted on the basis of the individual *S*'s arousal change during presentation of a stimulus, and is independent of the association value or other quality unique to a stimulus. Th finding provides further support for th theory of reverberating neural circuit previously used (Kleinsmith & Kaplan 1963) to explain both the marked reminiscence effect in high arousal learning and the more typical forgetting curve of low arousal learning.

In brief, high arousal learning shows poor immediate recall due to the relative unavailability of the rapidly reverberating neural trace, and strong permanent memory due to the great amount of perseverative consolidation. Low arousal learning, on the other hand, shows good immediate recall due to lack of rapid reverberation, and yet poor long-term memory for the same reason.

REFERENCES

Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural, and medical research.* London: Oliver & Boyd, 1938.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

Kleinsmith, L. J., & Kaplan, S. Paired-associate learning as a function of arousal and interpolated interval. *J. exp. Psychol.*, 1963, **65**, 190–193.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Lykken, D. T. Properties of electrodes used in electrodermal measurement. *J. comp. physiol. Psychol.*, 1959, **52**, 629–644.

(Received March 1, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 127–131

SPATIAL AND TEMPORAL UNCERTAINTY AS DETERMINANTS OF VIGILANCE BEHAVIOR[1]

JACK A. ADAMS AND LAWRENCE R. BOULTER
University of Illinois

An experiment was conducted to evaluate the effects of temporal and spatial uncertainty on visual monitoring behavior in a complex vigilance task with spatially arranged stimulus sources. The 3 stimulus sources of the display were arrayed over 144°. A 2 × 2 randomized factorial design was used, with temporal certainty and uncertainty about when a signal will occur as one dimension of the design, and spatial certainty and uncertainty about where a signal will occur as the other dimension. All Ss had a practice and a criterion session of 2.5 hr. duration, and there were 20 Ss in each of the 4 groups. Mean response latency for speed of signal detection increased as temporal and spatial uncertainty increased. Results were discussed in terms of temporal expectancies that cue S about the time of signal occurrence, and the observing response which directs S's head and eyes for the reception of visual stimuli.

The characteristics of human attention have been most frequently studied in the laboratory within the context of vigilance situations where prolonged monitoring of displays is required. A critical signal class is defined, and *S*'s task is to be alert and report the occurrences of signals. Most vigilance tasks are simple, with a single source to be monitored (Frankmann & Adams, 1962), but the special problems and issues for complex visual tasks with multiple, spatially arrayed sources for emitting signals have been receiving emphasis in recent research (Adams & Boulter, 1960, 1962; Adams & Humes, 1963; Adams, Humes, & Stenson, 1962; Adams, Stenson, & Humes, 1961). Two interesting properties of complex visual tasks such as these, which importantly determine when and where *S* will look to receive the next signal, are temporal and spatial uncertainty. Temporal uncertainty is defined by the temporal regularity of signals, and it is assumed to determine temporal expectancies, or a state of readiness in *S* for when the next signal will occur. Temporal expectancy can be considered a learned state that is acquired through experience with the temporal intervals in the task, and it is inferred from performance as a function of the time intervals between signals. Spatial uncertainty is defined by the regularity with which signals occur at the spatially arrayed stimulus sources. As with temporal expectancies, *S* can learn spatial expectancies about where the next signal will occur through experience in the task, and this learning can be expected to direct the visual observing responses that he employs for stimulus reception (Wyckoff, 1952). Frankmann and Adams (1962), in their review of the vigilance literature, discuss hypotheses and methodology that govern contemporary research on spatial expectancies. Our particular method for controlling observing responses was to manipulate the amount of spatial uncertainty, which in turn was related to the predictability that *S* could acquire about the place of signal occurrence. With spatial certainty, *S* could quickly learn where the next signal would occur, and only a minor observing response was required to be properly oriented for the next signal. But with high spatial uncertainty, where spatial patterning of signals was random, *S* had to maintain an active scanning of stimulus sources to insure speedy signal detection because he never could be sure where the next signal occurrence would take place. Thus, through learning, a measure of control over the observing response was gained.

---

[1] This research was supported by Contract AF 19(604)-5705 monitored by the Operational Applications Laboratory, Deputy for Technology, Electronic Systems Division, Air Force Systems Command. The assistance of Herbert H. Stenson in the collection and analysis of the data is gratefully acknowledged.

Using a complex vigilance task with spatially arrayed stimulus sources, and a long monitoring session, the experiment reported here investigated the separate and combined effects of temporal and spatial uncertainty of signal occurrence on human attentive behavior.

METHOD

Apparatus.—Two units of the Multiple Source Vigilance Apparatus were used, each in a separate room, and two *S*s were run at a time from a central control station operated by *E* in an adjacent building. The room had general ceiling illumination with a 100-w. bulb, and *S* sat at a desk and faced a semicircular metal track fixed at eye level to the back of the desk. On the track, and 30 in. from *S*'s eyes, were mounted three small ($1\frac{1}{2} \times 2$ in. front surface) digital display boxes that presented the number 20 ($\frac{5}{16}$ in. high) which was the signal that *S* had to detect. The number was bright enough for easy reading, but the brightness level was intentionally low so that *S* could not catch the signal in peripheral vision. He had to orient to a display box each time. The three display boxes were arrayed over 144°, with one box placed in the center position directly in front of *S* and the other two boxes placed 72° to the left and right of center. The 144° required head and eye movements to scan the boxes, and this wide separation was used intentionally to elicit an ample, distinctive observing response.

The boxes were normally blank. When a signal appeared *S*'s task was to report it as quickly as possible by pressing a large button located 4 in. from his finger tips. He was instructed always to respond with the same, preferred hand. The latency between onset of a signal and *S*'s response to it was the performance measure, and it was recorded on a Standard Electric timer at *E*'s station. A signal remained on for 5 sec., and the signals and the intervals between them were programed on punched tape and automatically read by a digital tape reader.

The *S*'s chair was locked into a position that was judged by *S* to be comfortable but which inhibited him from moving the chair or leaving it. Furthermore, *S*'s arms were kept in a standardized position by requiring *S* to keep them on padded armrests on the desk top, which was slanted at an angle of 20° for comfort. The armrests were wired so that *E* was informed by telltale lights whenever *S* moved his arms from their required position or if he used his nonpreferred hand, and *E* would then issue a reminder over an intercom system. One or two reminders during the practice session were sufficient to guarantee these procedures throughout the criterion session that followed. The main purpose of the armrests was to control the distance of arm-hand movement, which is a well-known variable for response time (e.g., Brown & Slater-Hammel, 1949).

Experimental procedures.—Each session was 2.5 hr., divided into five trials of 30 min. each for scoring purposes. A trial had nine signals. All *S*s had a practice and a criterion session, which were administered on separate days. The practice session was administered to insure that *S* was thoroughly familiar with the task and procedures, and to provide an opportunity for learning about the conditions of spatial and temporal uncertainty. Therefore, only the data for the criterion session were analyzed.

The experiment had an independent group of *S*s in each cell of a 2 × 2 factorial design with temporal certainty and uncertainty as one dimension, and spatial certainty and uncertainty as the other dimension. For *temporal uncertainty*, the following intersignal intervals were randomized separately on each of the five trials: 15, 33, 81, 129, 180, 234, 291, 354, and 438 sec. For *temporal certainty* the intersignal interval was constant at 195 sec., which was the mean value of the intervals for the temporal uncertainty condition. *Spatial uncertainty* had the nine signals separately randomized among the three sources on each trial, with the restriction that each of the three sources be assigned three signals on a trial. *Spatial certainty* was a repetitive movement of signals from left to right, right to left, etc., and again with three signals at each source on a trial. To help guarantee the random characteristics of signals and to minimize *S*'s learning of the particular characteristics of an input tape, a different tape with the same events in a newly randomized order was used for the practice and criterion sessions for the groups that had temporal and/or spatial uncertainty.

Each of the four groups had 20 *S*s. Group TU (temporal uncertainty) had temporal uncertainty and spatial certainty. Group SU (spatial uncertainty) had spatial uncertainty and temporal certainty. Group C (certainty) had both temporal and spatial certainty, and Group U (uncertainty) had both spatial and temporal uncertainty. Thus, Group C had signals that were perfectly predictable in time and place, Group TU had signals predictable only in place, Group SU had signals predictable only in time, and Group U had signals that were uncertain in both time and place. Instructions for a group specified the certain and uncertain features of the task, and this information supplemented the familiarization provided by the practice session.

Subjects.—The 80 *S*s were university male undergraduates who were paid for their participation. They were randomly assigned to groups. Their watches were taken from them at the start of each session.

RESULTS

Percent detection.—Previous research that has used tasks with persistent signals has always found a very high level of detection (Adams & Boulter, 1960, 1962; Adams & Humes, 1963; Adams et al., 1962; Adams et al., 1961; Boulter & Adams, 1963), and this experiment was no exception. All groups had a detection level of 98.5% or higher in the criterion session.

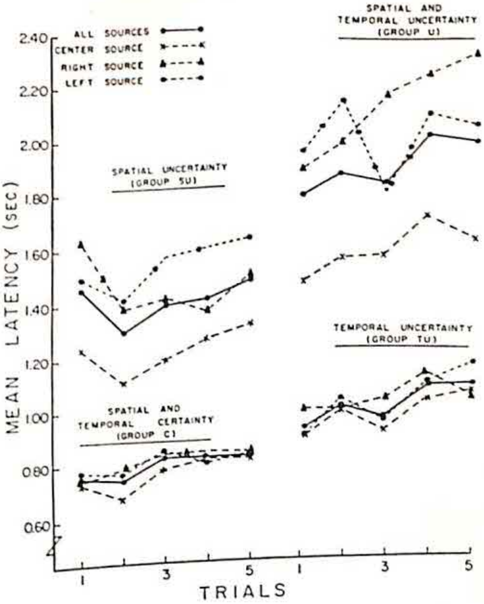


FIG. 1. Mean response latency as a function of trials for each of the four groups.

Latency measure.—The score for an *S* on a trial was the mean of his responses on the trial, which were almost always nine because of the high detection rate. Figure 1 shows a plot of mean latency for each group as a function of the five trials of the criterion session. Mean performance is shown for each of the three stimulus sources, as well as for overall performance where all three sources were combined. Group C, which had signals perfectly predictable in time and spatial position, had the best overall performance. Group TU, whose signals were only temporally uncertain, had somewhat poorer performance than Group C. Groups SU and U had performance distinctly poorer than the other two groups, showing that the requirement for extensive visual scanning when signals had high spatial uncertainty is impairing for vigilance performance. And, the compounding of spatial and temporal uncertainty for Group U induced performance that was considerably poorer than Group SU which just had spatial uncertainty. The grand means for Groups C, TU, SU, and U were 0.80, 1.01, 1.41, and 1.91 sec., respectively. Simple analysis of variance for between-groups differences, where the score for each *S* was his overall mean performance for the session, was statistically significant, $F(3, 76) = 66.56$, $p < .01$. The prominent group differences in Fig. 1 were substantiated statistically.

A three-way analysis of variance—Trials (5) × Sources (3) × *S*s (20)—with an *S*'s trial score being the entry, was performed for each group to test the effects of stimulus sources and trials. All groups in Fig. 1 show some tendency for performance to decrease as a function of trials, which is the well-known vigilance decrement, but only Group TU had a significant trials effect. For the trials variance, the F ratios (4, 76) and probabilities for Groups C, TU, SU, and U, respectively, were: 1.75, $p > .05$; 3.24, $.01 < p < .05$; 1.73, $p > .05$; 1.50, $p > .05$. Only Group SU and U, with high spatial uncertainty, had sources as a significant source of variation. For the sources variance, the F ratios (2, 38) and probabilities for Groups C, TU, SU, and U, respectively, were: 1.51, $p > .05$; 1.75, $p > .05$; 10.02, $p < .01$; 39.32, $p < .01$. No group had a Trials × Source interaction term that was significant at the .05 confidence level.

The standard deviation for each group was computed using the 300 trial scores of the session (20 *S*s, 5 trials, 3 sources), and their values were 3.08, 3.89, 5.03, and 7.27 sec. for Groups C, TU, SU, and U, respectively. Increasing temporal and spatial uncertainty not only worsened mean performance but also increased variability.

Discussion and Conclusions

Temporal and spatial uncertainty both emerge as clear-cut variables for vigilance behavior in a complex visual task with multiple stimulus sources. It was a reasonable expectation for Group C, with a relatively easy task of learning the repetitive temporal and spatial patterning of signals, to have a high level of proficiency after 2.5 hr. of learning in the practice session, but the outcomes for the other groups were not so self-evident. The amount of temporal uncertainty was the variable differentiating Groups C and TU and Groups SU and U, and in both instances it induced lower proficiency. Under conditions of high temporal uncertainty *S*s presumably had poor temporal expectancies and an unreliable internal mechanism for cuing the time of signal arrival, and the result was a reduced state of readiness to react and longer response latencies.

Groups C and TU had low demands placed on the observing response because their *S*s could easily learn in the practice session where the next signal would occur, and their performance levels were high relative to Groups SU and U where spatial uncertainty elicited an active observing response as *S* continually searched for a signal whose place of occurrence was unknown to him. With signal occurrence random with respect to the three spatial locations of stimulus sources, the observing response, being directional, had only one chance in three of being oriented to the correct stimulus source at the time of signal occurrence. The result was longer average latencies and relatively high variability. The observing response, as a small-scale response for stimulus reception that intervenes between stimuli and the instrumental motor response system whose action *E* records, has also been revealed as a potent determiner of performance in studies where amount of spatial separation of stimulus sources was a variable (Adams & Boulter, 1962; Adams & Xhignesse, 1960; Briggs & Howell, 1959; Fitts & Simon, 1952; King, 1961).

A possible explanation of vigilance decrement is that *S* gives up active visual scanning as the long session progresses. While recognizing that the direct recording of head and eye movements would be most decisive for evaluating this hypothesis, our data do allow a cautious inference about such matters. Decrement was not most prominently associated with Groups SU and U that had the greatest requirement for active observing, as might have been expected had these groups abandoned or reduced this demanding requirement after a while. Only Group TU had a statistically significant amount of decrement, and all other groups had a small but nonsignificant decremental trend over trials. However, other studies (Adams & Boulter, 1960; Adams & Humes, 1963; Adams et al., 1962) have suggested that the observing response is related to decrement where some stimulus sources are clustered with respect to others and S's observing appears to become preoccupied with the cluster.

The small amounts of decrement are consistent with our other research on complex tasks with multiple stimulus sources and persistent signals (Adams & Boulter, 1960, 1962; Adams & Humes, 1963; Adams et al., 1962; Adams et al., 1961). Although vigilance decrement is often substantial for very simple monitoring tasks with brief, transitory signals (Frankmann & Adams, 1962), the attentive powers of the human *S* in complex visual tasks appear to be far greater than many would surmise.

REFERENCES

Adams, J. A., & Boulter, L. R. Monitoring of complex visual displays: I. Effects of response complexity and intersignal interval on vigilant behavior when visual load is moderate. *USAF CCDD tech. Note*, 1960, No. 60-63.

Adams, J. A., & Boulter, L. R. An evaluation of the activationist hypothesis of human vigilance. *J. exp. Psychol.*, 1962, **64**, 495–504.

Adams, J. A., & Humes, J. M. Monitoring of complex visual displays: IV. Training for vigilance. *Hum. Factors*, 1963, **5**, 147–153.

Adams, J. A., Humes, J. M., & Stenson, H. H. Monitoring of complex visual displays: III. Effects of repeated sessions on human vigilance. *Hum. Factors*, 1962, **4**, 149–158.

Adams, J. A., Stenson, H. H., & Humes, J. M. Monitoring of complex visual displays: II. Effects of visual load and response complexity on human vigilance. *Hum. Factors*, 1961, **3**, 213–221.

Adams, J. A., & Xhignesse, L. V. Some determinants of two-dimensional visual tracking behavior. *J. exp. Psychol.*, 1960, **60**, 391–403.

Boulter, L. R., & Adams, J. A. Vigilance decrement, the expectancy hypothesis, and intersignal interval. *Canad. J. Psychol.*, 1963, **17**, 201–209.

Briggs, G. E., & Howell, W. C. On the relative importance of time sharing at central and peripheral levels. *USN Train. Devices Cent. tech. Rep.*, 1959, No. 508-2.

Brown, J. S., & Slater-Hammel, A. T. Discrete movements in the horizontal plane as a function of their length and direction. *J. exp. Psychol.*, 1949, **39**, 84–95.

Fitts, P. M., & Simon, C. W. Some relations between stimulus patterns and performance in a continuous dual-pursuit task. *J. exp. Psychol.*, 1952, **43**, 428–436.

Frankmann, J. P., & Adams, J. A. Theories of vigilance. *Psychol. Bull.*, 1962, **59**, 257–272.

King, W. J. Coordination in a complex tracking task as a joint function of spatial separation and predictability of stimuli. Unpublished master's thesis, University of Illinois, 1961.

Wyckoff, L. B., Jr. The role of observing responses in discrimination learning. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received March 2, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 132–141

COMPOUND STIMULI IN PAIRED-ASSOCIATE LEARNING[1]

LEONARD M. HOROWITZ, LOUIS G. LIPPMAN,[2] SANDRA A. NORMAN,
AND GEORGE W. McCONKIE

Stanford University

In several paired-associate (PA) tasks, triads of words served as stimuli. The associative relationship between the words of a triad was systematically varied. Words of each H (horizontal) triad were interassociated, while those of the I (independent) triads were not. Words of the V (vertical) triads were associated with words of other V triads but not with each other. In Experiment I, 30 Ss learned each type of item. Performance on the H items was facilitated, and that on the V items was hindered ($p < .001$). Experiment II ($N = 30$) replicated this result and provided control data with single-word stimuli. PA performance with stimulus triads was always poorer than that with single-word stimuli ($p < .001$). Experiment III ($N = 60$) examined the generality of these results. A single view of PA learning is proposed to integrate several interpretations of the results, and implications for serial learning are discussed.

This article reports three experiments that examine the effect of compound stimuli on paired-associate (PA) learning. Two of these experiments show that PA learning is facilitated when the ingredients of the compound stimulus are interassociated. The third experiment, however, limits the generality of this conclusion. These experiments provide empirical data that may help clarify the nature of the stimulus in serial learning.

The "chaining hypothesis" of serial learning asserts that each item of a serial list comes to elicit the following item, but a number of direct experimental tests (e.g., Primoff, 1938; Young, 1959) have disconfirmed this hypothesis. In these experiments a serial list and a PA list were prepared such that each pair of adjacent items of the serial list became a pair of associates in the PA list; *S* then learned both lists in either order. Hereafter, this PA task is called a "chained PA task."

According to the chaining hypothesis, positive transfer should occur between tasks, but in general, research has shown little such transfer. Horowitz and Izawa (1963) demonstrated positive transfer when the internal interference in the serial list was minimized. Even so, their results suggested that the chaining hypothesis is oversimplified. It implies, for example, that the two tasks be learned with equal speed, but in actual fact, serial learning was mastered sooner.

To account for their results, Horowitz and Izawa (1963) proposed the following "cluster hypothesis," which views serial learning as a more complex PA task. When *S* learns a serial list with little internal interference, a *cluster* of preceding items serves as the effective stimulus for each item. To account for their results with this hypothesis, though, the authors had to make one major assumption—that the complex PA task is facilitated if the elements of the stimulus cluster are themselves interassociated. Then, when *S* learns a serial sequence (A-B-C-D-E), Item E becomes a response, say, to the preceding cluster BCD. During serial learning, the elements of the cluster become inter-

---

[1] This research was supported in part by Grant MH 05428-02 of the United States Public Health Service.

[2] Now at Michigan State University.

associated (Horowitz & Izawa, 1963), and by the above assumption, the complex association is facilitated. Thus, this assumption allowed the authors to explain how serial learning might be mastered sooner than the chained PA task.

A first step in examining the cluster hypothesis, then, is to test this assumption. In Exp. I and II, *S* learned associations between stimulus triads and single responses; the associative relationship between stimulus words was systematically varied. The experimental arrangement also permitted the authors to compare PA learning with single and compound stimuli. In Exp. III, the authors tried to broaden the generality of their results with a modified procedure.

EXPERIMENTS I AND II

Method

Materials.—Table 1 presents the lists of stimulus-response pairs that were used in Exp. I and II. Each stimulus is composed of a triad of words, and the relationship between these words was varied to make three types of triads. The words of each H (horizontal) triad were associates of one another in the Connecticut norms (Bousfield, Cohen, Whitmarsh, & Kincaid, 1961); insofar as possible they were not associated with words of other triads. Those of the V (vertical) triads were associated with words of other V triads but not with each other. Words of the I (independent) triads were not associated with one another nor with words of other triads.

The responses of List 1 were selected from Noble's (1952) list of dissyllables. The mean association value of the H responses was 1.27; of the V responses, 1.36; of the I responses, 1.30. The responses of List 2 were selected from Melton and Safier's list of two-syllable adjectives (Hilgard, 1951). They were so selected that each response of List 2 occurred more than 10 times per million in the Thorndike-Lorge (1944) frequency count, and no two of them were synonyms.

Procedure, Exp. I.—First *S* pronounced each response of List 1 once, and the nine triad-response pairs were then presented on a Lafayette memory drum. Each stimulus triad was exposed as a unit for 2 sec. and then together with its response for 2 sec. On successive trials *S* tried to anticipate the response by pronouncing it. Four different orders of presentation were adopted; the procedure was continued to a criterion of two perfect trials. The instructions encouraged *S* to attend to the entire triad (as he must during serial learning) and told him of a later test of stimulus recall.

Some procedure was needed to demonstrate that the functional stimulus for each response was in fact more than a single word. Therefore, *S*s were divided into three groups, and each group participated in a transfer task. New triad combinations were formed of the 27 stimulus words in Table 1, and the same 9 responses were paired with them. (The authors intentionally retained the same pool of 27 stimulus words to avoid irrelevant

TABLE 1

PA LISTS USED IN EXP. I AND II

H Items		V Items		I Items	
List 1—Exp. I					
THIRSTY WATER DRINK	POLEF	BITTER ROBBER CARPET	VOLVAP	SHEEP HEALTH TABLE	SAGROLE
KNIFE FORK SPOON	QUIPSON	SWEET THIEF RUG	GOJEY	SING LAW NEEDLE	WELKIN
TOBACCO SMOKE CIGARETTE	LATUK	SOUR STEAL FLOOR	RENNET	DREAM SPEED CHURCH	NEGLAN
List 2—Exp. II					
SUN MOON STARS	AWFUL	SOFT HEALTH FOOT	ANGRY	COLLAR HOCKEY PART	SLEEPY
LAKE RIVER STREAM	POLITE	HARD DOCTOR HAND	UNSEEN	WALTZ PAIL GRIEF	OPEN
PETAL FLOWER ROSE	EMPTY	SMOOTH SICKNESS LEG	EAGER	CHILL HAMMER PIG	SECURE
THIEF ROBBER STEAL	PROFOUND	FATHER MOUSE CHURCH	EVEN	SAUCER JUSTICE BOX	OUTWARD
CITY VILLAGE TOWN	FAITHFUL	CHILD RAT RELIGION	FANCY	LIQUOR EIGHT STOVE	SIMPLE
BLUE RED GREEN	EARLY	MOTHER CAT BIBLE	DISTINCT	CORK SPEAR KEY	COSTLY

influences due to new stimulus words.) One word of each original triad continued to be paired with its response of original learning, but the other two stimulus words were changed. If a single word served as functional stimulus during original learning, transfer performance should have been best when that word remained intact; performance should have been inferior, though, when either nonfunctional word remained intact, and equally so for each of them. Any other pattern of relative performances would imply that the functional stimulus contained more than a single word.

For one transfer group the first word of the H stimuli was held intact, and the remaining stimulus words were changed. For example, THIRSTY, FORK, and CIGARETTE were paired with POLEF to form a new pair. The second word of the V stimuli was held intact, and the third word of the I stimuli was held intact. For another transfer group the first word of the V stimuli, the second word of the I stimuli, and the third word of the H stimuli were held intact. For a third transfer group the first word of the I stimuli, the second word of the H stimuli, and the third word of the V stimuli were held intact. The procedure for learning the transfer task was identical to that of original learning, but *S* continued to a criterion of only one perfect trial. Finally, *S* was supplied with a listing of the responses, and he was given unlimited time to record the corresponding stimulus triads of original learning.

Procedure, Exp. II.—Experiment II served two purposes. First, it replicated the results of Exp. I and extended the generality of them. Second, it provided control data to demonstrate that the results of Exp. I were not artifacts of uncontrolled item characteristics.

As a practice task, *S* first learned a modified version of List 1. In this list each pair of associates contained a single stimulus word, the first word of the triad shown in Table 1. In all other respects, *S*'s task duplicated the original task of Exp. I. Thus, the authors expected similar performances on the H and I items (but inferior performance on the V items because of associative interference).

After *Ss* achieved a criterion of one perfect trial, they were divided into two groups. One group learned the 18 S-R pairs of List 2, and the other group, an alternate form of List 2. (In the alternate form, triads and responses were differently paired, and the order of words within triads was changed. The performance of these two subgroups did not differ systematically, and they are not distinguished hereafter.) The procedure for both subgroups was identical to that of Exp. I, but because List 2 was so long, some *Ss* did not complete the task. Therefore, learning was continued to a criterion of one perfect trial or to the end of the 50-min. session, whichever occurred first. Then *S* was supplied with a listing of the responses, and he was given unlimited time to record the corresponding stimulus triads.

Subjects.—Thirty *Ss* participated in each experiment. These *Ss* were randomly selected from the introductory psychology classes at Stanford University. Some of them had learned CVC trigrams in earlier experiments, but this experience was considered unimportant in view of the random selection of *Ss*.

Results

Original learning (Exp. I and II).—Figure 1 shows the mean number of correct anticipations that *S* produced throughout the first 15 trials on List 1. Separate graphs are presented for the H, I, and V items, and each graph approaches 3.00, the maximum score. An analysis of variance was performed on these data and showed that the mean performances differed significantly, $F\ (2, 58) = 35.13$, $p < .001$. The difference between *S*'s performance on the H and I items approximately equaled the difference between his performance on the I and V items; $t = 0.53$ for the mean difference between the difference scores throughout 15 trials.

The importance of the stimulus composition was further illustrated by two other performance measures—the total number of errors that *S* produced before mastering the list and the trial number when *S* produced his last error. Both measures showed significant differences ($p < .001$) among the three types of stimulus composition. For the error measure, $F\ (2, 58) = 32.59$; for the trial measure, $F\ (2, 58) = 27.93$.

The control data from the practice task of Exp. II are also presented in Fig. 1. Each stimulus in this task

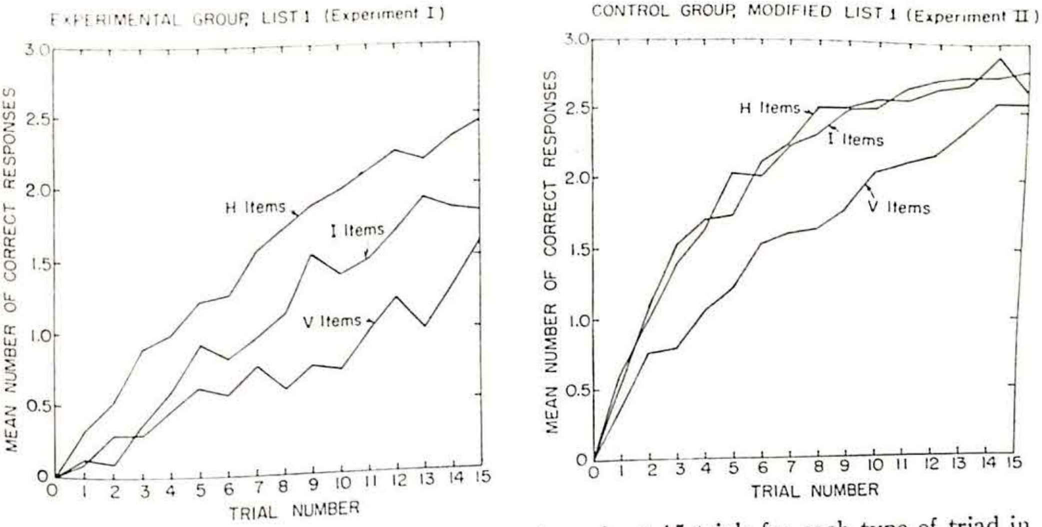


FIG. 1. Mean number of correct responses throughout 15 trials for each type of triad in List 1 (Exp. I); also, the corresponding control performances when single words served as stimuli (Exp. II).

contained just one word, and the H control pairs did not differ from the I control pairs ($t = 0.04$). Thus, the differences observed in Exp. I are not artifacts of uncontrolled item characteristics. The mean performance on the V control pairs is also presented in Fig. 1 for comparison. The V performance was significantly poorer; an analysis of variance showed significant differences among the three classes of control items, $F\,(2, 58) = 16.95$, $p < .001$.

The H and I control performances consistently surpassed the H experimental performance. In fact, the H experimental performance resembled the V control performance and did not differ significantly from it. (The value of t between the mean number of correct anticipations throughout 15 trials on the H experimental and V control items was 0.24.) The authors consider this result fortuitous, but it does underscore an important interference component in the H condition.

Several features of the experimental conditions help identify this interference component. For one thing, more stimulus words were available to interfere with the correct responses. Second, *S* probably used part of his learning time associating ingredients of the triads. And finally, the anticipation period may have been appreciably shortened because *S* had more material to read whenever a triad appeared (Schulz & Tucker, 1962).

Replication (Exp. II).—The major result of Exp. I was replicated in Exp. II as shown in Fig. 2. The *S*'s

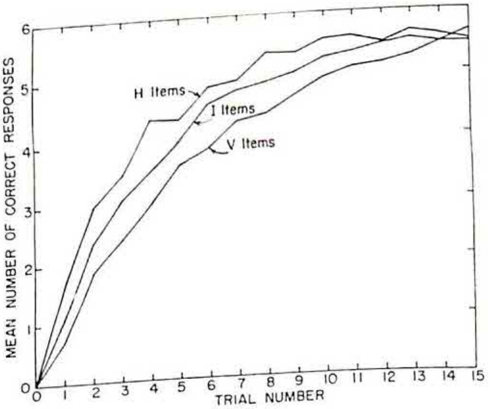


FIG. 2. Mean number of correct responses throughout 15 trials for each type of triad in List 2 (Exp. II).

performance on the II items consistently surpassed that on the I items, and his performance on the V items was consistently the poorest; $F(2, 58) = 15.72$, $p < .001$. Analyses were also performed of the error measure and of the trial measure described above. These analyses confirmed the significance of the difference reported in Fig. 2. For the error measure, $F(2, 58) = 13.78$, $p < .001$; for the trial measure, $F(2, 58) = 16.83$, $p < .001$.

Thus, these results confirm the authors' original hypothesis. If a response is learned to a compound stimulus, PA performance is facilitated when the compound's elements are interassociated.

When elements of different triads were associated, PA performance was severely hindered. If the cluster hypothesis is valid, this result is relevant to serial learning. In serial learning successive clusters contain elements in common, and such clusters would resemble the V triads of the present study. Thus, any theory of serial learning based upon the cluster hypothesis would have to explain how *S* overcomes the resulting interference.

The explanation might incorporate the following feature of serial learning. An element common to several clusters of a serial task occurs in a different position of each cluster, and its associative strength with the response may correspondingly vary. Thus, the strength of a cluster (as a unit) to elicit a given response may depend upon the *relative* strengths of the individual elements, and the interference in such PA learning could be thereby reduced. The analysis of scores on the transfer task showed that the position of an element within the triad does influence its associative strength with the response.

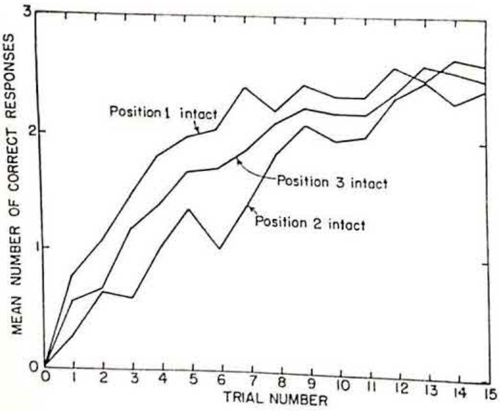


FIG. 3. Mean number of correct responses throughout 15 trials of the transfer task for each position of the original triad held intact.

Transfer task (Exp. I).—The experimental design of the transfer task required Latin square combinations of three variables: G (the three groups of *S*s), P (the position of the stimulus element held intact), and T (the nature of the stimulus triad). An analysis of variance was performed on the number of correct anticipations during the first 15 trials. Variable T did not affect transfer performance significantly, $F(2, 54) = 1.14$, $p > .10$, but Variable P did, $F(2, 54) = 10.52$, $p < .001$. If the first element remained intact, performance was best; if the second element remained intact, performance was poorest. The component of the $P \times T$ interaction variance that is associated with systematic group differences was significant, $F(2, 27) = 4.14$, $p < .05$, but the remaining component was not, $F(2, 54) < 1.00$.

Figure 3 presents the mean scores on each trial of the transfer task for each position held intact. Thus, the first element of the triad was associated most strongly with the response during original learning, and the second element, least strongly. This result is ambiguous, of course, since the inference is based on averages. Therefore, individual records were examined; the relative performances of 16 *S*s resembled that described in Fig. 3. No other single rank ordering systematically described the relative performances of the other 14 *S*s. At least for the average *S*, then, the functional stimulus of original learning contained more than one word.

Stimulus recall (Exp. I and II).—An analysis of the original stimulus words that *S* could recall confirmed the importance of the item's triad position during original learning. In Exp. I, the mean number of words in the first position that *S* recalled correctly was 6.10; in the second position, 3.77; in the third position, 4.87. These means differed significantly, $F\ (2, 232) = 14.78$, $p < .001$. (The error term in this analysis was the interaction, *S*s × P, T combinations.) The corresponding means of Exp. II were 12.60, 11.57, and 12.63; $F\ (2, 232) = 3.70$, $p < .05$. In both experiments, significantly fewer words of the second position were recalled, but the superiority of the first element in recall did not persist.

The nature of the stimulus triad also influenced *S*'s performance on stimulus recall. In Exp. I a mean of 6.83 H elements was correctly recalled; V elements, 3.77; and I elements, 4.10. The corresponding means of Exp. II were 15.37, 9.93, and 11.50. In both experiments the difference among means was significant, $p < .001$. In Exp. I, $F\ (2, 232) = 30.67$; in Exp. II, $F\ (2, 232) = 67.13$. The *S*'s recall of H words was understandably superior in view of the facilitative effect of interitem associative strength on free recall (Deese, 1959).

Analysis of errors (Exp. II).—Table 2 presents an analysis of the errors produced in Exp. II throughout the first 10 learning trials for each type of item. The errors have been classified as "failures to respond," "extralist intrusions," and "within-list errors." In addition to the actual

TABLE 2

Analysis of Errors throughout 10 Trials of Learning in Exp. II and III

Exp. II ($N = 30$)	H Items *n*	H Items %	I Items *n*	I Items %	V Items *n*	V Items %
Failures to respond	401	68.6	594	78.9	647	71.8
Extralist intrusions	10	1.7	11	1.5	25	2.8
Within-list errors	174	29.7	148	19.7	229	25.4
H responses	92	15.7	30	4.0	36	4.0
I responses	54	9.2	89	11.8	29	3.2
V responses	28	4.8	29	3.9	164	18.2
Total	585		753		901	
Exp. III ($N = 60$)	H Items *n*	H Items %	I Items *n*	I Items %	C Items *n*	C Items %
Failures to respond	957	68.1	1,029	73.9	839	72.9
Extralist intrusions	75	5.3	88	6.3	67	5.8
Within-list errors	374	26.6	276	19.8	245	21.3
H responses	131	9.3	74	5.3	68	5.9
I responses	140	10.0	107	7.7	60	5.2
C responses	103	7.3	95	6.8	117	10.2
Total	1,406		1,393		1,151	

omissions, responses were scored as failures if they occurred after the anticipation period. Extralist intrusions refer to all inappropriate responses, even stimulus words produced as responses. (On the whole, intrusions occurred very infrequently, but 43% of them were stimulus words.) Within-list errors refer to all the appropriate responses that were produced at the wrong time; they were further sorted into H, V, and I categories, according to the category of the response produced.

The most salient error was the failure to respond, which comprised more than 60% of the total errors. (For all pairs of comparisons the proportion of failures was significantly the larger, all z's $\geqslant 9.38$, $p < .001$.) This preponderance of failures is far more characteristic of a serial task than of a chained PA task (Horowitz & Izawa, 1963). Thus, the error pattern induced by the present PA task seems to resemble that of a serial task. Table 2 also shows how appropriate *S*'s within-list errors were. Within-list errors to H triads were typically H responses; to V triads, V responses; and to I triads, I responses. The value of χ^2 for a contingency table relating error type and item type was 214.73, $p < .001$. Before *S* mastered each pair of associates, he had apparently discriminated among the three classes of stimuli and had categorized the responses accordingly.

EXPERIMENT III

One major problem in examining the cluster hypothesis is that characteristics of serial learning are difficult to simulate with a PA task. During a serial task, for example, *S* obviously attends to all the items that precede a given item. But in the PA task of Exp. I and II, the authors had to encourage *S* specifically to attend to all ingredients of the stimulus. In Exp. III the authors examined a procedure that was designed to avoid these unusual instructions and still replicate the results of the earlier experiments. The words of each stimulus were presented singly and in sequence just as they would appear in a serial task. With this procedure the authors expected *S* to develop associative strength between each stimulus element and the response. Experiment III also provided a more direct comparison (a within-*S*s comparison) of PA learning with triads and with single words as stimuli.

Method

The materials of List 2 (Table 1) were modified as follows. Single-word control stimuli were prepared from the V triads by deleting two words of each triad. The words retained were not associates, and each one was paired with its response of List 2 to form six control (C) pairs. These C pairs were pooled with the H and I pairs of List 2 to form an 18-item list. Two sets of asterisks were added to each C stimulus to maintain a three-unit stimulus, so the total stimulus duration was uniform for all pairs. For half the *S*s, the asterisks of the C stimuli preceded the stimulus word; for the other *S*s, the asterisks followed the stimulus word. (This variable did not affect performance significantly, nor did its interaction with other variables, all *F*s ≤ 1.19, and it is not discussed further.)

Each pair of associates thus contained six ingredients which appeared singly and in succession on a Lafayette memory drum—the three stimulus elements, a blank space, the response, and then another blank space to separate pairs of associates. Two exposure rates were adopted for generality; for one group of *S*s each ingredient lasted 1 sec., and for the remaining *S*s, 2 sec. The *S* was not directly encouraged to attend to all stimulus words, nor was he ever tested for stimulus recall. Otherwise, the procedure was identical to that of Exp. II. Sixty *S*s were tested, all from the population of the earlier experiments.

Results

The total number of correct anticipations that S produced throughout seven trials was examined by a complete analysis of variance. This analysis showed that the 2-sec. exposure rate was significantly superior to the 1-sec. rate, $F\ (1, 56) = 15.57$, $p < .001$. Also, the nature of the stimulus triad exerted a significant influence over S's performance, $F\ (2, 112) = 14.98$, $p < .001$. No other sources of variation were significant.

Since there were no significant interactions, the data of all 60 Ss have been pooled in Fig. 4 to illustrate the effect of the stimulus composition. Figure 4 shows that the significance of the obtained difference was entirely due to the superiority of the single-word stimuli. Thus, one result of Exp. I and II was confirmed—a compound stimulus contributes interference beyond that present with single stimuli. However, the procedure of Exp. III did not reproduce the difference between the H and I items that was observed in the earlier experiments. If no differences had occurred at all in Fig. 4, the authors would have inferred that S had selected one word as stimulus from each triad. However, the resemblance between the H and I performances cannot be so explained.

The errors were examined in detail to learn why Exp. III failed to replicate the major result. This error analysis is presented in Table 2 along with the corresponding data of Exp. II. The failure to respond continued to be the most frequent error, but the errors of Exp. III differed in two major respects. First, the percentage of extralist intrusions exceeded the corresponding percentages of Exp. II (all z's $\geqslant 3.26$, $p < .01$). Thirty-six percent of these intrusions were stimulus words, and many of the remaining ones were associates of stimulus words. The stimulus triads seem to have been so poorly integrated that S could not easily separate stimuli and responses. Second, the within-list errors did not follow the clear pattern found in Exp. II. Within-list errors to C stimuli were typically other C responses, but the pattern of H and I errors could not be distinguished in this way. This result implies that the procedure of Exp. III did not allow S to perceive the difference between the H and I triads.

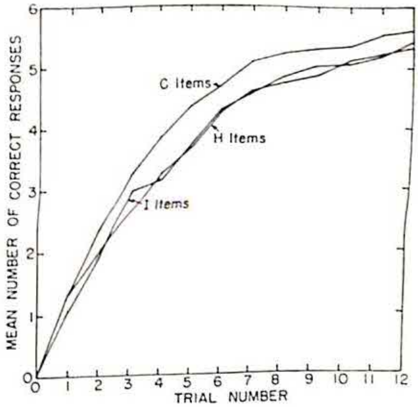


FIG. 4. Mean number of correct responses throughout 12 trials for each type of item in Exp. III.

DISCUSSION

Experiments I and II illustrate two major features of PA learning with compound stimuli. First, facilitation occurs when the elements of the stimulus are interassociated. Second, S does not perform as well with compound stimuli as he does with single stimuli. Both results have potential implications for serial learning. The advantage at the beginning of a serial list may occur partly because fewer items precede the responses (i.e., the stimulus compounds are smaller). Furthermore, serial learning may be simpler than the chained PA task because facilitation occurs as the items preceding a response become interassociated. Although the chained PA task has the advantage of one-word stimuli, these stimuli also occur as responses and interfere with *S*'s performance (Umemoto & Hilgard, 1961; Young, 1961).

The facilitation of the H items in Exp. I and II can be explained in three ways. First, stimulus integration may be a required stage of PA learning, requiring that *S* integrate the stimulus elements into a unit before he associates the unit with the response. This hypothesis can explain why PA learning is more difficult with compound stimuli and why PA performance is better with H items. Second, *S* may have an advantage learning the pairs that he can rehearse most easily during learning. The more readily *S* can recall a stimulus compound during learning, the more readily he can rehearse the S-R association. Elements of the H triads were more easily recalled during stimulus recall than those of the I triads, and this superiority probably occurs throughout learning. Finally, the task may require *S* to find some single associate that mediates an association between each stimulus element and the response. If the elements are interassociated, the probability increases that they will share some common associate and that one such mediator can be found.

These three explanations could be integrated by a single view of *S*'s task. When *S* is asked to learn an S-R pair, he may have to discover some verbal context which embodies the stimulus and the response. (Some *S*s have reported that they do structure their task in this way, though the context need not be well-verbalized by all *S*s.) The *S* could discover such a context for the H items more easily since the probability is greater (*a*) that these items are associates of some common item and (*b*) that they all elicit some common response. Then when *S* is presented with the stimulus words during testing, he recalls the entire context to produce the "correct response."

The statement that *S* must discover a mediator between stimuli and response is then operationally the same as the statement that he must integrate the stimuli. (This view assumes—with Asch & Ebenholtz, 1962—that a distinction between stimulus and response is unimportant for highly available items.) Furthermore, the more readily *S* can recall at least part of the context, the more easily he can rehearse the association during learning. Finally, the larger the stimulus compound, the more unusual the required context, and the more difficult *S*'s learning task.

Thus, associates that *S* learns rapidly should be ones that he easily incorporates into a single context. And he should be able to incorporate H items into contexts more easily than I items. To confirm this claim, 16 additional *S*s were asked to compose verbal contexts for the H and the I items of List 2. Each triad-response pair was presented singly on a 3 × 5 in. index card. The *S* composed a sentence which incorporated the four words, and he wrote his sentence on the blank page of a booklet. The order of items was systematically varied from *S* to *S*, and *S*'s latency was recorded. Latency is defined as the amount of time between an item's presentation and *S*'s first pencil stroke. The latencies were converted to logarithms, and mean log latencies were computed for each *S* on the H and I items separately. The mean log latency per H item was 1.016 log sec.; per I item, 1.244 log sec. ($t = 95.26$, $p < .001$). This difference was consistent throughout the data of all *S*s but one.

Whitman and Garner (1962) have presented results on a free recall task which resemble the results of the present study. They prepared three sets of geometric figures which varied along four different dimensions. The sets were alike, except for the form of the "internal structure": In the figures of Set A, there was no correlation among dimensions; in Set B, the levels of two dimensions were perfectly correlated; and in Set C, the levels of three dimensions were perfectly correlated. The *S*'s free recall was best for Set C and poorest for Set A.

Apparently, the correlation between dimensions of Set C permitted associative strength to develop between these dimensions. In this sense, the items of Set C resembled the H items of the present study. Within Set A, on the other hand, a level of any one dimension occurred with (and became associated with) several levels of the other dimensions; thus, Set A resembled the V items of the present study.

REFERENCES

Asch, S. E., & Ebenholtz, S. M. The principle of associative symmetry. *Proc. Amer. Phil. Soc.*, 1962, **106**, 135–163.

Bousfield, W. A., Cohen, B. H., Whitmarsh, G. A., & Kincaid, W. D. The Connecticut free associational norms. Technical Report No. 35, 1961, Contract Nonr-631(00), Office of Naval Research.

Deese, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

Horowitz, L., & Izawa, C. Comparison of serial and paired-associate learning. *J. exp. Psychol.*, 1963, **65**, 352–361.

Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

Primoff, E. Backward and forward associations as an organizing act in serial and in paired-associate learning. *J. Psychol.*, 1938, **5**, 375–395.

Schulz, R., & Tucker, I. F. Supplementary report: Stimulus familiarization in paired-associate learning. *J. exp. Psychol.*, 1962, **64**, 549–550.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

Umemoto, T., & Hilgard, E. R. Paired-associates learning as a function of similarity: Common stimulus and response items within the list. *J. exp. Psychol.*, 1961, **62**, 97–104.

Whitman, J. R., & Garner, W. R. Free recall learning of visual figures as a function of form of internal structure. *J. exp. Psychol.*, 1962, **64**, 558–564.

Young, R. K. A comparison of two methods of learning serial associations. *Amer. J. Psychol.*, 1959, **72**, 554–559.

Young, R. K. Paired-associate learning when the same items occur as stimuli and responses. *J. exp. Psychol.*, 1961, **61**, 315–318.

(Received March 4, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 142–147

EFFECT OF BODY TILT ON APPARENT VERTICALITY, APPARENT BODY POSITION, AND THEIR RELATION[1]

MARTIN BAUERMEISTER

Clark University

100 Ss indicated, by means of a luminescent rod (a) the vertical direction in space (apparent vertical); and (b) the direction of their longitudinal body axis (apparent body positions) under body tilt ranging from 90° left, through upright, to 90° right. Significant ($p < .001$) deviations of apparent from objective positions were observed for (a) apparent vertical, (b) apparent body position, and (c) the relation between the two. The trend of the observed deviations appeared as a nonlinear function of the degree of body tilt. The results were interpreted in terms of the notion of differentiation of a spatial reference system defined with respect to external space and a spatial reference system defined with respect to one's own body.

Since the time of Aubert (1860) the problem of the effects of body tilt on the perception of verticality and its relation to the perception of the degree of tilt of one's own body has been attacked frequently. However, systematic experiments utilizing a wide range of body tilts and assessing both apparent verticality and apparent body position by objective measurements have not as yet been conducted and are presented in this study.

METHOD

Apparatus.—The *S* was tilted in a heavy iron framework which was pivoted on a supporting frame so that it could be rotated 360°. Body and head of *S* were fully supported by adjustable boards. A luminescent rod, 1 m. long, 3 cm. wide, served as indicator of apparent verticality and apparent body position. The rod was pivoted on a supporting frame so that it could be rotated 360°. The pivots of the supporting frames of tilting board and luminescent rod were 3 m. apart and at the same height. The *S* was adjusted so that his center of gravity was lined up sagitally with the two pivots. The experiment took place in a windowless room, the walls of which were painted black. The room was separated by a black partition through which the shaft of the luminescent rod was led, so that only the rod was exposed to *S*. When the light was switched off no contours of the room were visible.

Subjects.—One hundred *S*s, 55 male, 45 female, were tested. Their ages ranged from 16 to 26; the average age was 18 yr.

Test conditions.—Each *S* was tested under 20 conditions of body tilt: 90° Left, 80° L, 70° L, 60° L, 50° L, 40° L, 30° L, 20° L, 10° L, −0°, +0°, 10° Right, 20°R, 30°R, 40° R, 50° R, 60° R, 70° R, 80° R, 90° Right. The nine positions of body tilt to the right were systematically randomized. Each tilt to the right was then followed by an equally large tilt to the left. Two upright positions were used. One of these 0° positions was measured after a tilt to the left side and is represented as "+" 0°, the other 0° position was measured after a tilt to the right side and is represented as "−" 0°. The two 0° positions were distributed randomly between the positions of tilt.

The stimulus object was presented by the adjustment method in the form of serial exploration, i.e., *E* adjusted the stimulus object according to the order of *S*. The *S* was instructed to tell *E* in which direction to move the luminescent rod and where to arrest it so that it looked (*a*) "straight up and down" (vertical), and (*b*) "in line with *S*'s body" (coinciding in direction with *S*'s longitudinal body axis). A test trial was given to ensure that *S* understood the instructions precisely. In each of the 20 body positions each *S* made two judgments on verticality and two judgments on body position. For the first judgment the luminescent rod was started 40° to the right (clockwise) of objective vertical (objective body position), for the second judgment 40° to the left (counterclockwise) as viewed by *S*. Fifty *Ss* started with judgments on apparent vertical, the remaining 50 *Ss* with judgments on apparent body position. The room was darkened during the time in which *S* was tilted, made his judgments, and was tilted back again. Between the tilts *S* stood erect in the lighted room. The rod was exposed during the time of each of the four judgments. Between the judgments the room was completely dark.

---

[1] This investigation was supported in part by Public Health Service Research Grant, MH-00348-13, from the National Institute of Mental Health, and conducted while the author was a North Atlantic Treaty Organization Research Fellow at Clark University. It is part of a PhD dissertation, supervised by S. Wapner and H. Werner. This work was done in part at the Computation Center at Massachusetts Institute of Technology, Cambridge, Massachusetts.

Measures.—Deviations of apparent from objective vertical were measured with respect to the plumbline. Deviations of apparent from objective body position were measured with respect to the objective degree of tilt. Clockwise (CW) deviations, as viewed by *S*, were designated as "+," counterclockwise (CCW) deviations as "−."

Treatment of data.—The raw data were prepared for statistical analysis by computing points of subjective equality for apparent vertical and apparent body position. The point of subjective equality was determined as the midpoint between the positions in which the stimulus object appeared vertical (parallel to *S*'s body) when the rod was started to the left and when it was started to the right of objective positions. A Treatment × *Ss* analysis was applied to test the overall effect of body tilt on apparent vertical and on apparent body position. The variance between the treatments (tilt) was further analyzed with respect to polynomial trend components (to the fifth degree) according to the trend analysis suggested by Grant (1956) to show the trend of deviations of apparent from objective positions as a function of the degree of body tilt. Grant's analysis was further utilized to compare deviations of apparent vertical with deviations of apparent body position.

RESULTS

Apparent vertical.—Increasing body tilt led to changing deviations of apparent from objective vertical. These changes appeared as a nonlinear function of the degree of tilt. They are presented in Fig. 1. With upright body position apparent vertical deviated to the left (CCW) of objective vertical. With body tilt up to 10° right (CW) and left (CCW) there was a tendency to displace apparent from objective vertical in the direction of body tilt. With further increasing body tilt up to 50° right (CW) and 40° left (CCW) this tendency reversed so that there was an increasing displacement of apparent from objective vertical opposite to the direction of body tilt. With further increasing body tilt up to 90° right (CW) and left (CCW) the tendency reversed a second time so that once again there was an increasing displacement of apparent from objective vertical in the direction of body tilt. The two reversals of the displacement tendencies were confirmed by the statistical trend analysis presented in Table 1. In addition to significant overall effects of body tilt the analysis showed for the trend of the total curve of deviations between 90° of body tilt to the left (CCW), through upright, to 90° of body tilt to the right (CW) significant quintic, cubic, and linear polynomial trend components, which characterize the total curve as having four significant inflections.

Apparent body position.—Increasing body tilt led to changing deviations of apparent from objective body positions. These changes appeared as a nonlinear function of the degree of tilt. They are presented in Fig. 1. With upright body position the apparent body position deviated to the left (CCW). With body tilt up to 20° right (CW) and 10° left (CCW) there was a tendency to displace the apparent body position in the direction of body tilt. With further increasing tilt up to 30° right (CW) and 50° left (CCW) the tendency reversed so that there was an increasing displacement of the apparent body position opposite

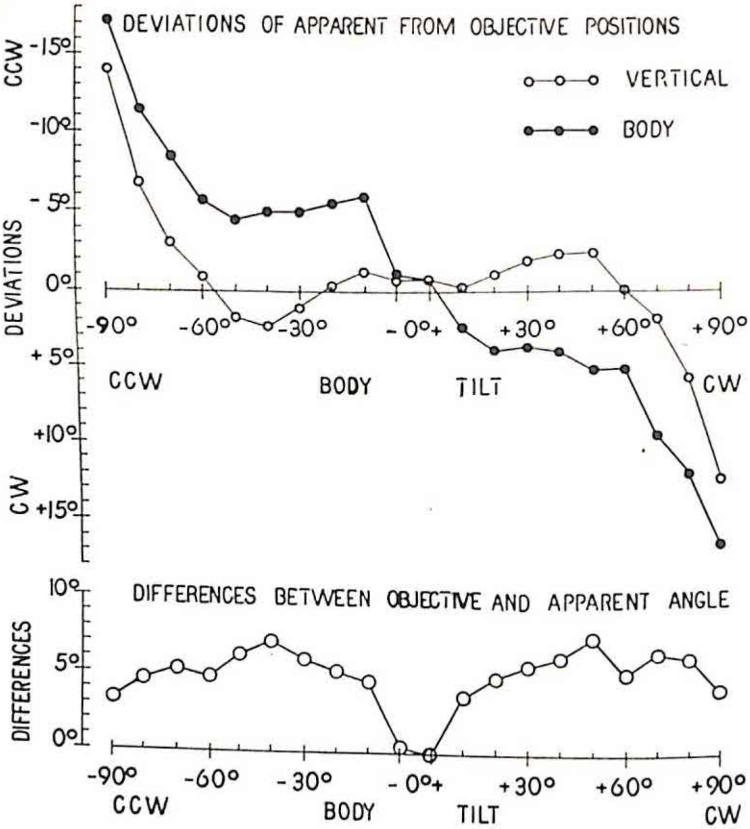


FIG. 1. Deviations of apparent from objective vertical and of apparent from objective body position under body tilt (upper part) and differences between apparent and objective angle of tilt (apparent angle always larger than objective angle) (lower part).

TABLE 1

EFFECTS OF BODY TILT ON APPARENT VERTICAL AND APPARENT BODY POSITION

Source of Variation	*df*	Vertical		Body Position	
		MS	*F*	*MS*	*F*
Subjects (*S*)	99	42.6		98.9	
Treatments (T)	19	2,414.4	78.3***	6,628.9	95.1***
T × *S*	1,881	30.8		69.7	
Polynomial trend components					
Linear	1	17,462.2	67.6***	116,408.1	221.0***
Quadratic	1	1.6	0.0	264.9	6.3
Cubic	1	26,332.9	362.0***	3,547.7	20.2***
Quartic	1	53.1	2.9	260.3	7.4
Quintic	1	1,267.9	39.0***	3,738.6	62.0***
Error terms for polynomials					
Linear	99	257.8		526.7	
Quadratic	99	29.9		41.9	
Cubic	99	72.7		175.4	
Quartic	99	18.4		35.0	
Quintic	99	32.5		60.3	
Total	1,999	54.1		133.5	

*** $p < .001$.

to the direction of body tilt. With further increasing body tilt up to 90° right (CW) and left (CCW) the tendency reversed a second time so that once again there was an increasing displacement of the apparent body position in the direction of body tilt. The two reversals of the displacement tendencies were confirmed by the statistical trend analysis presented in Table 1. In addition to significant overall effects of body tilt the analysis showed for the trend of the total curve of deviations between 90° of body tilt to the left (CCW), through upright, to 90° of body tilt to the right (CW) significant quintic, cubic, and linear polynomial trend components, which characterize the total curve as having four significant inflections.

Relation between apparent vertical and apparent body position.—There were similarities and differences in the deviations of apparent from objective positions of apparent vertical as compared to apparent body position as the body was tilted increasingly. The important similarity pertains to the reversal of the tendency of displacement for both, apparent vertical and apparent body position. Displacement in the direction of body tilt reversed to displacement opposite to the direction of tilt and reversed a second time to displacement in the direction of tilt. Accordingly the total curves of deviations for both vertical and body position show four inflections as confirmed by significant quintic polynomial trend components. The difference between the deviations of apparent vertical as compared to apparent body position pertains to the degree of displacement as a function of the degree of body tilt. For all positions of tilt the apparent body position deviated more in the direction of tilt than did the apparent vertical; furthermore, increasing body tilt did not lead to entirely identical quantitative changes in the degree of deviation. This resulted in differences in details of shape of the two curves. The curve representing the deviations of apparent from objective vertical appears as an **S** shaped curve with two additional minor inflections, whereas the deviations of apparent from objective body position fit a sloping curve with four inflections. These differences are confirmed statistically. Table 2 shows a significant Treatment (tilt) × Task (vertical vs. body) interaction and indicates significant differences for linear (slope) and cubic (**S** shape) polynomial trend components for the two curves.

Another way of expressing the relation between apparent vertical and apparent body position is in terms of the angle between their positions, which will be termed "apparent angle." Increasing body tilt led to nonlinear changes in the relation between apparent and objective angle of tilt, which are presented in the lower part of Fig. 1. The apparent angle was for all positions of body tilt larger than the objective angle. Further, with increasing body tilt up to 50° right (CW) and 40° left (CCW) there was an increase of the apparent angle which reversed to a decrease with body tilt increasing beyond 50°.

DISCUSSION

There are three main findings: (*a*) The absolute values of the deviations of apparent from objective positions of vertical and longitudinal body axis were relatively small and oscillated around the objective positions; (*b*) the positions of apparent vertical and apparent longitudinal body axis changed as a nonlinear (quintic) function of the degree of body tilt; and (*c*) the apparent angles (angle between apparent vertical and apparent body position) between the

TABLE 2

SIGNIFICANCE OF DIFFERENCES BETWEEN DEVIATIONS OF APPARENT VERTICAL VS. BODY POSITION

Source of Variation	*df*	*MS*	*F*
Between *S*s	199	73.2	
Between tasks (vertical vs. body)	1	17.2	0.2
Error between	198	73.4	
Within *S*s	3,800	95.3	
Between treatments (tilt)	19	7,453.0	145.8***
Polynomial trend components			
Linear	1	110,623.0	278.5***
Quadratic	1	119.0	3.3
Cubic	1	23,819.4	210.8***
Quartic	1	264.0	10.1***
Quintic	1	4,435.4	89.1***
Treatments × Task	19	1,478.7	28.9***
Polynomial trend components			
Linear	1	20,991.5	52.8***
Quadratic	1	216.4	5.9
Cubic	1	5,931.8	52.5***
Quartic	1	35.2	1.3
Quintic	1	246.3	4.9
Error within	3,762	51.1	
Linear	198	397.2	
Quadratic	198	36.4	
Cubic	198	113.0	
Quartic	198	26.3	
Quintic	198	49.8	
Total	3,999	94.2	

*** $p < .001$.

vertical and the longitudinal body axis were always larger than the objective angles and changed with increasing body tilt, such that the apparent angle was minimal with body erect, increased to a maximum at 40°–50° tilt, and decreased with further body tilt.

The *S* was obviously able to locate a stimulus object with respect to two spatial reference systems: (*a*) the reference system of external space, as evidenced by use of the luminescent rod indicating the position of one of its main coordinates, the vertical; (*b*) the reference system of one's own body, as evidenced by use of the luminescent rod indicating the position of one of its main coordinates, the longitudinal body axis.

In keeping with "sensory-tonic field theory of perception" as developed by Werner and Wapner (1949, 1952, 1956; Wapner & Werner, 1957) it is assumed that a perceptual property corresponds to a particular relation between organismic state (psychophysiological) and stimuli issuing from an object (psychophysical); this proposition applies to the reference systems of external space as well as body space. Variations of body tilt led to changes of the organismic state, which pertain to changes of the perceptual properties under consideration, i.e., the positions of the main coordinates of the two spatial reference systems. The positions of these coordinates change systematically with body tilt: with change from erect to small and medium body tilts (60°) the position of apparent vertical is increasingly opposite the direction of body tilt; with change from erect to small body tilts (20°) the apparent body position is increasingly in the direction of body tilt followed by relatively little change until 60° body tilt; with larger angles of body tilt (beyond 60°) there is an increase in shift of both coordinate systems in the direction of body tilt. The interplay of these seemingly opposite reaction tendencies (in the direction of tilt vs. opposite the direction of tilt) may be interpreted in terms of a dynamic organization of the two spatial reference systems in terms of polar organismic tendencies to "go with" the stimulation or "counteract" it. (For a more extensive discussion of this interpretation see Bauermeister, 1962.)

The distinctiveness between spatial reference systems defined with respect to external space and defined with respect to one's own body may be regarded as the outcome of a developmental process of differentiation between the two. The angle between the positions of the main coordinates of each of the two systems may be considered as an indicator of the degree to which the two spatial reference systems have become differentiated (cf. McFarland, Clarkson, Wapner, & Werner, 1961), which mirrors the degree of articulation between organism and environment. The theoretical significance of the relation between the degree of differentiation of the two reference systems depending on the degree of body tilt is more comprehensively discussed elsewhere (Bauermeister, 1962).

REFERENCES

Aubert, H. Eine scheinbare bedeutende Drehung von Objekten bei Neigung des Kopfes nach rechts oder links. *Virchows Arch. pathol. Anat. Physiol. klin. Med.*, 1860, **20**, 381–393.

Bauermeister, M. The relation between subjective body space and objective external space under conditions of body tilt. Unpublished doctoral dissertation, Clark University, 1962.

Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–153.

McFarland, J. H., Clarkson, F. E., Wapner, S., & Werner, H. Relation between perceptual properties of objects in space and one's own body. Paper read at Eastern Psychological Association, Philadelphia, April 1961.

Wapner, S., & Werner, H. *Perceptual development.* Worcester, Mass.: Clark Univer. Press, 1957.

Werner, H., & Wapner, S. Sensory-tonic field theory of perception. *J. Pers.*, 1949, **18**, 88–107.

Werner, H., & Wapner, S. Toward a general theory of perception. *Psychol. Rev.*, 1952, **59**, 324–338.

Werner, H., & Wapner, S. Sensory-tonic field theory of perception: Basic concepts and experiments. *Riv. psicol.*, 1956, **50**, 315–337.

(Received March 5, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 148–150

DECAY OF PRISM AFTEREFFECTS[1]

CHARLES R. HAMILTON AND JOSEPH BOSSOM[2]

California Institute of Technology

8 Ss practiced reaching movements for 15 min. while viewing their hands through laterally oriented 20-diopter wedge prisms. Following adaptation, Ss sat quietly in the dark for an additional 15 min., after which prism adaptation was found to be significantly reduced ($p < .005$). This result indicates that under these conditions the normal visuomotor coordinations are not lost and replaced by an equivalent set of prism-altered ones, as suggested by previous work, but are in fact retained and reinstated following prism removal in the absence of any visual feedback.

When wedge prisms, with their bases oriented left or right, are placed before the eyes of human *S*s, the immediate effect is a lateral deflection of the visual field in the direction of the prism's apex, with consequent errors of localizing and aiming responses. As shown by many investigators, these prism-induced errors became quickly corrected with the practice of visually guided movements (Smith & Smith, 1962). After one has adapted to prisms, their removal leads to similar errors in the opposite direction, called aftereffects. These aftereffects in turn undergo rapid correction and normal coordination is restored. It has often been proposed that adaptation to prisms reflects the same processes used in the establishment or maintenance of normal visuomotor integration, and that therefore there is no basic difference between normal and prism-altered coordinations (Held & Bossom, 1961; Kohler, 1951; Snyder & Pronko, 1952). If in fact the normal and prism-altered coordinations do represent alternative states of the same system, then it would be expected that the same conditions would be required for the return of normal coordination after prism adaptation as are necessary for the establishment of prism-altered ones.

In recent work on human *S*s it has been shown that self-produced movement and the entailed visual feedback, termed reafference after von Holst (1954), are necessary for prism adaptation (Held, 1961). In the following experiment the necessity of visual reafferent stimulation for the return of normal coordination following prism adaptation is investigated.

METHOD

The apparatus used for adapting and testing *S* is similar to one described by Held and Gottlieb (1958); the procedure employed is presented in more detail elsewhere (Hamilton, 1964). Briefly, *S* adapted his reaching movements to a laterally displaced visual field by moving his arm for 15 min. while watching this motion through 20-diopter wedge prisms. The tests measured accuracy of reaching to the virtual image of a target without prisms and without visual feedback of the movement. The difference between the pre- and postadaptation tests gave the magnitude of the aftereffect.

The survival of the aftereffect error was then checked on each *S* under two contrasting conditions. In the "reafferent condition" visual feedback was intentionally used to help restore normal coordination; *S* moved his hand back and forth for 15 min. while watching the movement without the prisms in the apparatus. In the "nonreafferent condition" visual feedback was completely absent; *S* sat quietly in the dark for 15 min. At the end of each of these periods the aftereffect was again determined.

---

[1] This research was carried out with the aid of United States Public Health Service predoctoral and postdoctoral fellowships awarded to the senior and junior authors, and partially supported by United States Public Health Service Grant M3372 awarded to R. W. Sperry. The authors wish to thank R. W. Sperry particularly for his help in preparing the manuscript.

[2] Now at National Institutes of Health, Bethesda, Maryland.

Eight young adults were randomly assigned in the following way for testing. Four *Ss* viewed the preferred and four the nonpreferred hand during adaptation. Half of each group was run on the nonreafferent condition first while the other half began with the reafferent condition. The aftereffect was re-established and the second condition was tested at least 3 days after the first. All *Ss* had a right-hand preference and all were adapted with the bases of the prisms to *S*'s left.

RESULTS

The average reductions in the aftereffect errors found after 15 min. of normal reafferent exposure and after 15 min. in the dark are presented in Table 1. All values in this table are given in tenths of an inch; 1 in. equaled approximately 3°. These values are the means of the individual differences between the magnitudes of the aftereffects before and after the test conditions. Both the decrease of the aftereffect with the reafferent condition, $t\ (7) = 3.59, p < .005$, and the decrease with the nonreafferent condition, $t\ (7) = 4.15, p < .005$, are significantly greater than zero. The aftereffect decrease for the situation with visual feedback is not significantly greater than that for the dark condition in this experiment, $t\ (14) = 1.16, p < .2$. The average aftereffect before decay was 10.7 units for each group.

TABLE 1

DECREASE IN AFTEREFFECTS WITH AND WITHOUT VISUAL REAFFERENCE

Cond.	*N*	*M*[a]	*SD*
Reafferent	8	8.6	6.8
Nonreafferent	8	5.4	3.7

[a] Values reported in .1-in. units.

DISCUSSION

Since *Ss* lose the prism aftereffects not only when they can view their own movements, as expected, but also when they sit quietly in the dark, it may be inferred that visual reafferent stimulation is not necessary for the return to normal visuomotor coordination following prism adaptation. A similar decay has been observed incidentally in the course of other experiments dealing with intermanual transfer of prism aftereffects, where instead of sitting in the dark *S* was actively engaged in adapting his other hand to the same prism displacement (Hamilton, 1964). In this case the aftereffects of the hand first exposed decayed even while the second was being adapted. These data are also consistent with a previous observation on split-brain monkeys in which a rather rapid decrement in the magnitude of the aftereffects occurred following prism removal even though there was little opportunity for visually guided movement (Bossom & Hamilton, 1963).

These decay experiments show that loss of prism aftereffects is not exactly equivalent to elimination of prism-induced errors and brings into question the earlier assumption that there are no basic differences in the neural processes underlying normal and prism-altered coordinations. Additional experiments would be desirable to determine whether this decay reflects (*a*) a superiority, due to greater practice, of the old coordination over the newly established ones, (*b*) the operation of innate mechanisms that favor normal visuomotor coordination, or (*c*) the ability of tactile or proprioceptive reafference, which was not specifically controlled in this experiment, to reinstate normal coordination following prism adaptation.

REFERENCES

BOSSOM, J., & HAMILTON, C. R. Interocular transfer of prism-altered coordinations in split-brain monkeys. *J. comp. physiol. Psychol.*, 1963, **56**, 769–774.

HAMILTON, C. R. Intermanual transfer of prism adaptation. *Amer. J. Psychol.*, 1964, in press.

HELD, R. Exposure history as a factor in maintaining stability of perception and coordination. *J. nerv. ment. Dis.*, 1961, **132**, 26–32.

HELD, R., & BOSSOM, J. Neonatal deprivation and adult rearrangement: Complementary techniques for analyzing plastic sensory-motor coordinations. *J. comp. physiol. Psychol.*, 1961, **54**, 33–37.

HELD, R., & GOTTLIEB, N. Technique for studying adaptation to disarranged hand-eye coordination. *Percept. mot. Skills*, 1958, **8**, 83–86.

KOHLER, I. Über Aufbau und Wandlungen der Wahrnehmungswelt. (Trans. by G. Krauthamer) *S. B. öst. Akad. Wiss.*, 1951, **227**, 1–118.

SMITH, K. U., & SMITH, W. M. *Perception and motion.* Philadelphia: Saunders, 1962.

SNYDER, R. W., & PRONKO, N. H. *Vision with spatial inversion.* Wichita: Univer. Wichita Press, 1952.

VON HOLST, E. Relations between the central nervous system and the peripheral organs. *Brit. J. anim. Behav.*, 1954, **2**, 89–94.

(Received March 6, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 151–161

ASSOCIATION BY CONTIGUITY[1]

NORMAN E. SPEAR,[2] BRUCE R. EKSTRAND, AND BENTON J. UNDERWOOD

Northwestern University

3 experiments were designed to study the role of contiguity in establishing verbal associations. In Experiment I a transfer test followed learning of a paired-associate list which had compound stimuli (word and trigram). Experiment I revealed positive and negative transfer effects with appropriate and inappropriate pairing of the original stimulus components. Heavy stimulus selection was also apparent. Experiments II and III demonstrated that a pair of items in a discrimination task acquire a bidirectional association although instructions require only that they be discriminated. The data of all 3 experiments support the proposition that contiguous appearance of 2 verbal units will result in the development of an association between them while S is learning other associations.

The experiments reported here were designed to help answer the following question: If two verbal units are presented together, will an association develop between them even though *S* is not instructed to learn an association between them? Obviously, this question asks about the validity and fundamentality of the classical principle of association by contiguity. Modern work in verbal learning has, generally speaking, paid little attention to the mechanisms by which new associations are formed. The controversy between one-trial vs. incremental learning is concerned only with the characteristics of the product of the mechanisms of association, not with the mechanisms as such. When it is said that an association is established between two verbal units by mediation through a common term previously associated with each unit, we beg the question of how the associations became established in the first place.

If at sometime in the past contiguity had lost some of its appeal as an explanatory mechanism, there is scattered evidence that it is regaining status. Some examples: Spence (1956) has tentatively suggested that instrumental reward learning may be based solely on contiguity. Within the context of classical conditioning of cats, Brogden (1962) has recently concluded that ". . . contiguity is a sufficient condition of learning [p. 174]." A study of Osgood and Anderson (1957) could be interpreted as indicating that merely presenting two verbal units together results in the development of some associative connection between them. Indeed, one might interpret incidental verbal learning by a contiguity principle. However, most incidental learning studies have been based on acquisition of individual verbal units and it is difficult to identify the specific associative linkages. In the present studies a form of incidental learning is used, but the analysis is made of the associative connections between pairs of verbal units.

Two situations were used to assess the role of contiguity. In Exp. I two verbal units were used as a compound for each stimulus in a paired-

---

[1] This work was supported in part by Contract Nonr-1228 (15), Project NR 154-057, between Northwestern University and the Office of Naval Research.

[2] Now at Rutgers University.

associate list. Following learning of this list, *S* was given a second paired-associate list in which one of the elements of the compound was paired with the other. If an association develops as a result of the consistent contiguity of the two elements in the compound stimulus in the first list, positive transfer should result in learning the second paired-associate list. In Exp. II and III the contiguity was produced by verbal-discrimination learning in which pairs of items were presented, with *S* instructed to learn which member of each pair was the correct one. Subsequently, the pairs were used as stimulus and response terms in a paired-associate list. The learning of this latter list should be facilitated if an association developed between the two units during verbal-discrimination learning. This technique has been used by Battig, Williams, and Williams (1962) who reported no positive effect, a fact which will require later discussion.

EXPERIMENT I

In a recent study each compound stimulus in a paired-associate list consisted of a low-meaningful trigram and a frame of colored paper (Underwood, Ham, & Ekstrand, 1962). Transfer tests were given in which either the colored frames or the trigrams were used alone as stimuli in a second list in which the response terms were the same as in the first list and paired appropriately. Transfer was essentially 100% for the colored frames. For the trigrams, only a small amount of positive transfer was observed. The latter finding could result from a direct association between the trigram and the response term developed during first-list learning, or it could be due to mediation. Thus, if an association developed between the trigram and the color, the correct response on transfer tests could be mediated from trigram to color to response term, since the latter association is known to be strong. The critical question for the mediation interpretation is whether or not an association does develop between the elements of a stimulus compound, and to obtain data bearing on this question was the major purpose of Exp. I.

Method

General procedure.—The experiment had four conditions. Two of these were used merely to verify the results of the previous experiment, since three-letter words have been substituted for the color frames. The two groups learned an initial list of eight pairs with each compound stimulus consisting of a trigram and a word. On the transfer list, Group W had only the words as stimuli, Group T had only the trigrams as stimuli. On the basis of previous results, Group W would be expected to show essentially 100% transfer to the second list, and Group T to show low but significant positive transfer.

Group T-W, after learning the first list with the compound stimuli, was presented a list in which the trigrams were stimulus terms and the words were response terms. The pairings were appropriate; that is, the trigram and word which had appeared together as the compound stimulus in the first list were paired. If an association had developed between these two units, positive transfer would be expected in the learning of the second list. Group T-Wr also was given a second list in which the trigrams were stimuli and the words response terms, but the pairings were not appropriate, i.e., trigrams and words which had occurred together as the compound in the first list were *not* paired in the transfer list. If an association had developed between the two units of the compound stimulus in the first list, this re-pairing in the second list should produce negative transfer. Thus, if the performance of Group T-W on the second list is better than the performance of Group T-Wr, it may be taken as evidence that an association had developed between the two elements of the compound in learning the first list.

Detailed procedure.—All four groups of *Ss* learned exactly the same first list consisting of compound stimuli and single-digit numbers as response terms. The eight items were as follows: IGW, CAT-7; ZBN, NOT-6; DSU, BOY-9; TPM, DIE-3; OVJ, LEG-5; RZL, SEA-8; CFY, HIS-4; KHQ, RAN-2. The trigrams were of low interletter association value (Underwood & Schulz, 1960), and the words all have AA frequency ratings in Thorndike-Lorge (1944) except CAT, which has an A rating. In presenting the list, the two components of the stimulus were presented one above the other. There were four orders of the eight items; for each compound, the word appeared above the trigram on two orders, with the reverse positions for the other two orders. On any given order the trigrams were above the words for four pairs and below for the other four pairs. Anticipation learning was used throughout with the memory drum set at a 2:2-sec. rate.

The usual paired-associate learning instructions were given. In addition, *S* was fully informed concerning the nature of the stimulus terms. He was further informed that both the word and trigram would appear together consistently from trial to trial. The intent of the instructions was to inform *S* completely concerning the nature of the task without biasing him toward "using" one or the other stimulus components as the functional stimulus. Learning was carried to one perfect recitation of the pairs.

Approximately 60 sec. after learning the first list, the second or transfer list was presented. Again, *Ss* were fully informed concerning the nature of the transfer list. Groups T and W were instructed to give as many correct responses on the first presentation as possible. The *Ss* of Groups T-W and T-Wr were given one study trial on the transfer list before anticipation was requested. For all groups the learning of the second list proceeded for 10 trials.

Each of the four conditions contained an independent group of 24 *Ss*. The schedule sheet consisted of 24 blocks of 4 conditions each, each condition being represented in each block and each block consisting of one of the 24 different orders in which 4 conditions can be arranged. The 24 blocks were randomly ordered and *Ss* were assigned to conditions in order of their appearance at the laboratory. When an *S* was lost for any reason he was replaced by the next *S*.

Results and Discussion

All four groups had the same first list; the mean trials to learn varied from 10.42 to 12.83, ($F = .91$). It may be concluded that the four groups did not differ in learning ability.

Cue selection.—Group W had only the words as stimuli on the transfer list; Group T had only the trigrams. A comparison of the learning of these two groups on the transfer task will determine whether cue selection had occurred as was true in the previous study. On the first trial Group W had a mean of 6.88 correct responses, Group T had a mean of 1.92. Needless to say, the difference between these means is highly significant ($t = 8.96$). In the previous experiment (Underwood et al., 1962), a control group was used. The *Ss* in this group were given further trials on the first list after having reached a criterion of one perfect recitation on the seven-pair list. On the trial following the achievement of the criterion the performance of this group fell from 7 items correct to 5.5 items correct. In the present study Group W fell from 8 items correct to 6.88. The percentage drop is roughly equivalent for the two groups, indicating that for the present Group W the omission of the trigrams from the stimulus compound had little if any decremental effect on continued performance. This is to say that the words were completely effective as stimuli.

Group W maintained a high level of performance throughout the 10 transfer trials; Group T showed a slightly negatively accelerated learning curve but at the end of 10 trials was still below Group W. Group W gave a mean of 75.75 total correct responses on the 10 trials, Group T a mean of

52.08. The difference gives a t of 7.43. It is evident, therefore, that there was heavy selection of the words as functional stimuli in learning the first list. Yet, the fact remains that on the first presentation of the transfer list 1.92 items of the trigram-number pairs were given correctly by *S*s of Group T. Of the 24 *S*s, 10 got no items correct but 14 got one or more correct, with 1 *S* getting all items correct. On this first trial, a total of 192 responses was possible. A total of 52 errors (misplaced responses) occurred and 46 correct responses, or a mean of 4.08 responses of any kind. By chance (or guessing), therefore, roughly .5 items (out of the 8 possible) may have been given correctly. However, subtracting .5 from the 1.92 correct responses actually given leaves 1.42, which differs significantly from zero ($1.42 \pm .49$; $t = 2.90$). It is concluded that first-list learning positively influenced the learning of the trigram-number pairs in the second list.

Association between stimulus components.—If, as a result of first-list learning, associations developed between the two elements of the stimulus compounds, Group T-W, having trigrams as stimuli and words as responses for second-list learning, should show a positive effect. Group T-Wr, having the trigrams and words randomly paired, should show a negative effect. Since no neutral control group was used, the prediction is only that the performance on the transfer list should be better for Group T-W than for Group T-Wr. On the first anticipation trial of the transfer list, Group T-W gave a mean of 3.21 correct responses, Group T-Wr a mean of 2.00. The difference ($1.21 \pm .49$) produces a t ($df = 46$) of 2.47, which falls at about the 2% significance level. However, on total correct responses over the 10 trials the means are almost identical for the two groups.

The above facts, while not contradicting the notion that an association developed between the elements of the compound stimulus during learning of the first, certainly do not offer impressive evidence in support of the notion. This finding was especially puzzling in view of the error data. If an association does develop between the two members of the compound, interference should be present in learning the transfer list for Group T-Wr. Furthermore, this interference should be manifest in the errors, since it would be expected that the trigrams would elicit the words with which they had appeared in the compound stimulus of the first list. These responses, of course, are incorrect for the transfer list. Group T-Wr made a total of 165 overt errors in learning the transfer list of which 54 (33%) were "good" errors in the sense that they were words given to the trigram with which they had been paired in the first list. As will be more fully described shortly, a control group of 24 *S*s, not having learned the first list, learned the same transfer list as Group T-Wr. The errors made by these *S*s were scored in exactly the same manner as were the errors for Group T-Wr. This produced 16% "appropriate" errors. The proportion of such errors for each *S* was determined for both the control *S*s and those in Group T-Wr. Following an arc-sine transformation, the t was 2.36, significant at about the 2% significance level. Thus, the error data for Group T-Wr strongly support the notion that an association was developed between the two stimulus elements during first-list learning.

The discrepancy between the error measure and the correct-response measure suggested the possibility that the two transfer lists differed in difficulty. Both lists consisted of the same stimulus and response terms, but the random pairing of the items for Group T-Wr may have resulted in pairs which were more easily associated than were the pairs for Group T-W. To evaluate this possibility, two new groups of 24 *Ss* were used, one group being assigned the T-W list, the other the T-Wr list. Of course, these *Ss* were not given the first list; they simply learned the transfer lists for 10 trials.

On the first anticipation trial the mean number correct on the T-W list was 1.58; on the T-Wr list, 2.71. The mean difference (1.13 ± .34) gives a *t* of 3.34, indicating the greater difficulty of the T-W list. Another way to evaluate this is by a 2 × 2 analysis of variance in which the entries are the number correct on the first anticipation trial with the two lists as one variable and control-experimental group as the other. This analysis produced an *F* for interaction of 15.27, significant far beyond the 1% level. It seems therefore, that the T-Wr list was an intrinsically easier list to learn than was the T-W list. The fact that Group T-W was actually superior to Group T-Wr on the initial transfer trials of the main experiment in spite of the fact that the list was more difficult, gives support to the original presumption that an association developed between the elements of the stimulus compound during first-list learning. If this is true, the positive transfer observed in Group T *could* be due to mediation from trigram to word to number. Of course, while the evidence may be taken to indicate that an association was established between the two stimulus elements, this does not necessarily mean that the positive transfer in Group T occurred by the mediation noted above; the possibility of a direct association between the trigram and number has not been ruled out. This is to say that while a direct association between the two elements of the compound may have been developed by contiguity, it does not mean that the transfer observed in Group T was due to this; it may be due to a direct association between the trigram and number. Also, it is apparent that the associations which may have developed by contiguity were few, weak, or both. Finally, it should be noted that it is possible that evidence taken here to indicate learning by contiguity could be due to mediation. If a direct association had been established between the trigram and number during first-list learning, the transfer effects could have been produced by associations running from the trigram to number to word, the latter being a backward association. Although a subsequent study (not to be reported) made this seem unlikely, it could not be ruled out entirely. Therefore, Exp. II and III were performed in an effort to minimize mediation possibilities.

EXPERIMENT II

In verbal-discrimination learning, pairs of verbal units are presented to *S*. The task for *S* is to learn which unit of each pair is correct (as arbitrarily determined by *E*). In learning this task, both units in a pair are spatially and temporally contiguous on each presentation. Does an association develop between these two units? If so, subsequent pairing of them in a paired-associate list should produce positive transfer.

Likewise, pairing the units inappropriately on a subsequent paired-associate list should result in negative transfer.

Method

Lists.—Only two groups of *S*s were used. For one, the paired-associate list (PA list) had appropriate pairings (Group A), for the other, inappropriate (Group I). Both groups learned the *same* PA list; hence, appropriate and inappropriate pairings were determined by the pairings in the verbal-discrimination list (VD list). The units were low-frequency, three- and four-letter words. Low-frequency words were used to minimize pre-experimental interitem associative connections among the words. The 12 pairs in the PA list were as follows: BAIL-FIFE, RIG-JOT, DISC-LOAM, GENT-HUB, KEG-DOLT, JAB-TING, FAG-RUE, HALO-CARP, TAB-MUG, CAM-BOOR, MOOT-GLUT, LEER-KILT.

For Group A the VD list had the same pairings as above; thus, items paired in the VD list subsequently became items to be associated in the PA list. For Group I, the pairings in the VD list were inappropriate; for example, BAIL was paired with JOT, RIG with FIFE, and so on.

One unit of each pair was correct in VD learning, the other incorrect. Half the units which were correct in the VD list became response terms in the PA list, and the other half became stimulus terms. This is subsequently spoken of as a mixed list with regard to the correct–incorrect dichotomy. However, it is not a mixed list with regard to appropriate or inappropriate pairings in the PA list. That is, for Group A, all pairings were appropriate; for Group I, all pairings were inappropriate.

Procedure.—Four different orders of the pairs were used in presenting the VD list. The rate was 2:2 sec. During the first 2-sec. interval, the two units appeared together, one above the other, and *S* was to say aloud the correct member of the pair. During the final 2 sec., the correct member of the pair was shown alone. The *S*s guessed on the first presentation of the pairs. One unit occurred above the other on two of the four orders, and below the other on the other two, but this was nonsystematic across pairs.

The PA list also consisted of four different orders, and was also presented at a 2:2-sec. rate. The learning of the VD list was carried to two successive perfect recitations (*S* was instructed concerning the criterion), the learning of the PA list for 10 anticipation trials. However, before PA learning, *S*s in Group A were told that the two items were paired as they had been in the VD list and were asked to give any responses they could on the first presentation of the list. The *S*s in Group I were told that the words were the same as they had seen in the VD list but were *not* paired as they had been in the VD list. These *S*s were instructed not to anticipate responses until after they had seen the list once.

There were 20 *S*s in each group. Twenty blocks of two conditions each (I and A) were made up, with the order within each block (I-A, or A-I), being determined randomly. The *S*s were assigned to the schedule in order of appearance at the laboratory.

Results and Discussion

The VD list was learned rapidly. Group A required a mean of 5.40 trials, Group I, 6.35 trials. These values include the guessing trial and the two criterial trials. The difference between these two means is far from significant ($t = 1.10$).

The acquisition of the PA list over the 10 trials is shown in Fig. 1. It is apparent that Group A is superior to Group I. Across the 10 trials the mean total correct is 85.05 for Group A, and 59.80 for Group I. The difference (25.25 ± 6.12) gives a t of 4.13. It will be remembered that *S*s in Group A were asked to give any correct response they could on the first presentation of the PA list. The mean number correct was 3.15 responses. Of the 20 *S*s, only 3 failed to give a correct response; 2 *S*s gave seven correct responses. Thus, it would appear as if an association had developed between some of the pairs in VD learning. However, when all PA trials are considered, the learning of all pairs was not facilitated. For 10 of the 12 pairs, a very marked superiority is present for *S*s in Group

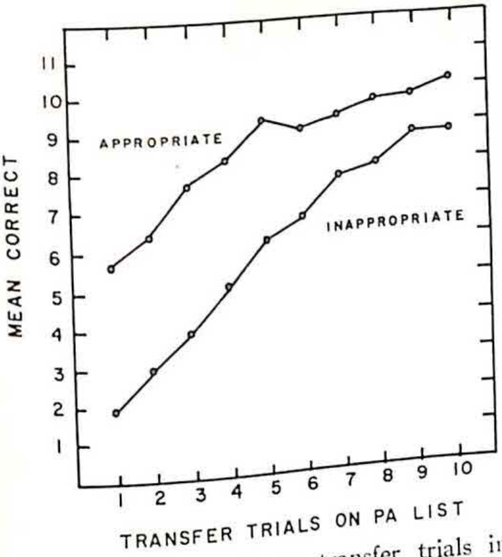


FIG. 1. Learning on transfer trials in Exp. II as a function of appropriate and inappropriate pairings of items previously given in a verbal-discrimination list.

A. For one pair, the two groups are essentially equal, and for the remaining pair, learning is appreciably faster for the I Group.

Across the 10 anticipation trials the mean number of overt errors was 14.7 and 28.5 for Groups A and I, respectively. Of greatest theoretical importance, however, is the fact that 20.0% of all overt errors made by the *S*s in Group I were "appropriate" errors. That is, the response given to the stimulus term had been paired with that stimulus term during VD learning. Thus, it seems probable that the difference between the two curves in Fig. 1 is a summation of two factors, namely, facilitation in Group A, and an interference in Group I. Both factors would be expected on the basis of the contiguity hypothesis.

The evidence of Exp. II is interpreted to mean that appreciable association by contiguity occurred in VD learning. Alternative mechanisms such as differentiation, familiarity, or response availability will not account for the differences between the two groups in PA learning. However, in view of the fact that Battig et al. (1962) had failed to find facilitation in going from a VD list to a PA list, further work seemed indicated.

EXPERIMENT III

Method

There were seven groups of 15 *S*s each. All groups learned the same PA list following VD learning. The PA list was the same one as used in Exp. II. Groups A and I were given conditions identical to those given the two groups in Exp. II. That is, following VD learning, six of the units which had been correct in VD learning became stimulus terms in PA learning, and six became response terms. In contrast, the four additional experimental groups described below had unmixed lists; that is, all the correct units in VD learning became stimulus terms or all became response terms.

Group A-S: Pairings in the PA list were appropriate; all units which were correct in VD learning became stimulus terms in the PA list.

Group A-R: Pairings were appropriate, but all units which were correct in VD learning became response terms in the PA list.

Group I-S: Same as Group A-S except pairings were inappropriate.

Group I-R: Same as Group A-R except pairings were inappropriate.

Group C: Control group which learned a VD list consisting of 12 pairs, none of which appeared in the PA list. The 12 pairs in the VD list had the same characteristics as those given *S*s in the other groups.

One change was made in presenting the VD list. After presenting the two units for 2 sec., during which time *S* was to make his choice between the two, both units were again presented for 2 sec. immediately, with the correct unit underlined. It was believed that this might make the learning somewhat more difficult, although in fact there is no evidence that it did.

All other details were the same as in Exp. II. A schedule was constructed of 15 blocks with each condition occurring once within each block, but with the order of the seven conditions being random within each block. The *S*s were assigned to the schedule in order of their appearance at the laboratory.

Results

VD learning.—The groups did not differ in rate of learning the VD lists. The seven means ranged from 5.47 to 7.80 trials, but the F (1.75) was far from significant. Assuming that the groups of *S*s did not differ in learning ability, the insignificant F may be taken to indicate that the seven different lists used did not differ in difficulty.

PA list learning.—The mean total correct responses given on the 10 trials of PA learning for each group are shown in Table 1. It should be noted that the means for all three A groups are greater than the mean for the C group, and that the means for all three I groups are lower than the mean for the C group. Thus, inappropriate pairing produced negative transfer, appropriate pairing, positive transfer.

TABLE 1

MEAN TOTAL CORRECT ON 10 TRIALS ON THE PAIRED-ASSOCIATE LIST IN EXP. III

Groups	*M*	*SD*
A	85.60	14.94
A-S	91.20	15.77
A-R	84.87	17.46
Overall	87.22	16.36
I	60.00	24.40
I-S	63.33	19.53
I-R	55.47	13.51
Overall	59.60	19.92
C	72.00	24.58

The performance of the three A groups did not differ appreciably (F = .65), and the performance of the three I groups also did not differ (F = .53). Thus, whether the correct responses in VD learning all become stimulus terms in PA learning, all become response terms in PA learning, or half and half, does not influence learning in either the A or I groups.

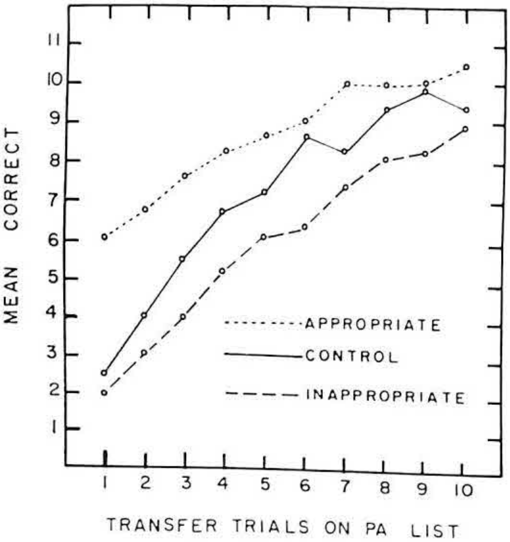


FIG. 2. Learning on transfer trials in Exp. III as a function of appropriate and inappropriate pairings as compared with a control group.

The three A groups were combined to give the upper curve in Fig. 2, and the three I groups were combined to give the lower curve. The overall means for the A groups (87.22) and for the I groups (59.60), are very nearly identical with the means obtained in Exp. II (85.05 and 59.80, respectively). Analysis of variance for the C group and the combined A and I groups gives an F of 22.52, which is significant far beyond the 1% level. The t between the C group mean and I group mean is 2.13, that between the C group and A group, 2.61. Both are significant beyond the 5% significance level.

It will be remembered that *S*s in the A groups, after being informed of the nature of the PA list, were asked to give any responses they thought they knew on the first presentation of the PA list. The 45 *S*s in these three groups averaged 4.04 correct responses on this trial. One *S* gave all 12 items correctly, and only one failed to give any correct responses. Thus, 33% of the items were given correctly on the first presentation of the PA list. That sheer guessing played a small role in this performance is suggested by the fact that an average of only 2.02 wrong responses was given.

PA item learning.—The total correct responses given to each item across all 10 trials was determined. These were summed for the three A groups and also for the three I groups, and then each sum for each item was divided by 3 to obtain the same base as for the C group. Comparing the A group scores with those of the C group showed that the values for the A group were greater for 11 of the 12 items. Testing whether the mean of this distribution of differences (19.17 ± 5.03) differs significantly from zero gave a t of 3.81. A comparable comparison of the C group and I groups shows the C group to be superior on 10 of the items, equal on 1, and inferior on 1. The mean of the distribution of differences (15.58 ± 4.04) is also significantly different from zero ($t = 3.86$). Thus, the data on item learning fully support the earlier analysis that negative transfer was produced in the I groups, positive transfer in the A groups.

Intrusion errors.—In Exp. II, 20.0% of the errors made by the I group in the PA list were "good" errors (intrusions); that is, they were response terms which had been paired with the stimulus terms in VD learning. The comparable percentages in the present experiment were 25.88%, 18.76%, and 24.35% for Groups I, I-R, and I-S, respectively. The errors made by the C group were scored in the same manner as for Group I; the percentage "intrusions" was 5.21%. Clearly, there is a strong relationship between negative transfer and intrusions in the present procedures.

Discussion

The transfer data of Exp. III, as well as that of Exp. II, indicate that considerable associative strength is developed between the two verbal units presented together in VD learning. It is inferred from these findings that the association was developed as a consequence of the two units appearing contiguously in VD learning. The associations which developed did so in an average of six presentations and they were completely bidirectional (either term could elicit the other).

Is there any other subtle factor which might account for the development of the associations in VD learning? There was reinforcement for the correct response in VD learning in the sense that *S* was shown the correct response following his overt anticipation. Is there some sort of spread of effect in the Thorndikian sense so that the two words somehow become associated via the reinforcement? This seems unlikely for several reasons. First, since an association between the two words was not, in the usual sense, formally practiced during VD learning, it is logically difficult to have this association reinforced. Second, when a spread of effect has been found (and it has often failed to appear) it has been of very small magnitude; it would not seem, therefore, that it could account for the rather great transfer which occurred in the present studies. Third, if a spread of effect is occurring, it should be directly related to the number of reinforcements given during VD learning. This is to say that the greater the number of correct responses for an item in VD learning, the greater should be the strength of the association. This expectation has been checked. The number of correct anticipations made for each pair for all *S*s for all A groups in VD learning was determined. This distribution was correlated with the number of times each item was given correctly on the first presentation of the PA list, since this measure provides the purest one of the effect of VD learning. The product-moment correlation was .48. With *df* = 10, an *r* of .58 is required for the 5% significance level. A second and more severe test was made. For each *S* in the A groups a determination was made of the three easiest and the three most difficult items in VD learning. Since more reinforcements occurred for the easier than for the difficult items in VD learning, if a spread of effect is responsible for the transfer noted the easier items should show the greatest transfer in PA learning. Direct-difference *t*'s between total correct in PA learning for the three easy and the three difficult items were calculated for each A group. For two of the groups (A and A-R) the direction of the difference favored the spread-of-effect interpretation, but the largest *t* (for Group A-R) was 1.70, with a *t* of 2.15 being required for the 5% significance level with *df* = 14. For Group A-S, the difference was in the opposite direction, but not significantly so (*t* = .26). Thus, all evidence considered, the spread-of-effect interpretation cannot be seriously considered as an alternative explanation for the present results.

As noted earlier, Battig et al. (1962) failed to find positive transfer in moving from a VD list to a PA list when appropriate pairing was used, although there was some evidence for negative transfer when inappropriate pairings occurred. Yet, it was clear that some associative connections were developed during VD learning. These investigators used nonsense syllables. Following VD learning, and immediately before PA learning, *S*s were presented one of the syllables from each pair, one at a time, and were asked to give the other syllable that was paired with it in VD learning. In one experiment, 17% correct responses were given, in another, 16%. Thus, some associations had developed between the pairs. The failure to find positive transfer in PA learning is probably due (these investigators suggest) to the fact that VD learning does not require response learning, and in PA learning with nonsense syllables much of the learning required is response learning or response integration. It is not entirely clear, however, why some positive transfer was not observable since some negative transfer was found with the inappropriate pairings.

Analysis of the Battig et al. results in conjunction with the present findings strongly suggests that meaningfulness interacts with learning by contiguity. The lower the meaningfulness the less the strength of the associations which occur as a consequence of contiguous presentations. The evidence for such a principle is as follows. First, there is the gross comparison between the use of nonsense syllables (in the Battig et al. study) and in the present Exp. II and III where words were used. Second, although the magnitude of the difference between appropriate and inappropriate pairings could not be assessed directly in Exp. I, it is clear that the effect was less than in Exp. II and III. This may be due to lower meaningfulness. In Exp. I, the compound stimulus, by which items were presented contiguously without *S* having an intent to learn them, consisted of a word and a very low-meaningful trigram. Finally, in Exp. III the most difficult PA items to learn (as gauged by the C group) were those least likely to be given on the first presentation of the PA list for the A groups. The product-moment correlation expressing this relationship was .59, which is just past the 5% significance level. On the assumption that the most difficult items to learn in the PA list are, in some manner, of lower meaningfulness than those which are easier to learn, a relationship between meaningfulness and amount learned by contiguity is seen.

To modify the principle of contiguity with this presumed interaction with meaningfulness will, of course, require additional explanatory mechanisms. One present guess is that less learning occurs by contiguity with low-meaningful items because such items are less likely to evoke an immediate and stable response (overtly or covertly) than are high-meaningful items. With low meaningfulness, on successive trials of VD learning "different" responses are appearing contiguously from trial to trial for the same pair, so that less learning of any specific association should occur. The notion that maximum learning by contiguity will occur when the verbal unit evokes a consistent and stable response is quite in line with interpretations of the role of meaningfulness in incidental learning (e.g., Postman, Adams, & Phillips, 1955).

In the strict sense, it cannot be concluded from the present experiments that contiguity is a sufficient condition for associative learning. The present operations always required *S* to learn something. Whether or not associative learning would occur if pairs were presented to *S* without requiring auxiliary learning is not handled by the present experiments.

It is concluded that associations may develop as a result of contiguous pairings of verbal units while these verbal units are being discriminated. No parallel conclusion is to be made that contiguity is the most efficient mechanism for learning. The usual college student *S* is equipped with many associations (associational aids) which he can transfer to a "new" learning task and perhaps supplement or displace learning by contiguity. Contiguity, however, may be used to explain how the associational aids, which make use of previous learning, were acquired in the first place.

REFERENCES

BATTIG, W. F., WILLIAMS, J. M., & WILLIAMS, J. G. Transfer from verbal-discrimination to paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 258–268.

BROGDEN, W. J. Contiguous conditioning. *J. exp. Psychol.*, 1962, **64**, 172–176.

OSGOOD, C. E., & ANDERSON, L. Certain relations among experienced contingencies, associative structure, and contingencies in encoded messages. *Amer. J. Psychol.*, 1957, **70**, 411–420.

POSTMAN, L., ADAMS, P. A., & PHILLIPS, L. W. Studies in incidental learning: II. The effects of association value and method of testing. *J. exp. Psychol.*, 1955, **49**, 1–10.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J., HAM, M., & EKSTRAND, B. Cue selection in paired-associate learning. *J. exp. Psychol.*, 1962, **64**, 405–409.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received March 6, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 162–167

PRIMARY STIMULUS GENERALIZATION UNDER DIFFERENT PERCENTAGES OF REINFORCEMENT IN EYELID CONDITIONING[1]

WILLIAM E. VANDAMENT AND LOUIS E. PRICE

University of Massachusetts

80 Ss were trained to a light of 3.20 apparent ft-c and tested to lights of less intensity in an eyelid conditioning investigation of primary stimulus generalization. All Ss were given 90 acquisition trials and 20 extinction trials, the 1st postacquisition trial serving as a test for pure generalization. ½ of the Ss received semirandom 50% reinforcement during acquisition. (a) The gradient of generalization for the 1st test trial displayed a systematic downward trend in response strength as a function of lessened magnitude from the CS. (b) The percentage of reinforcement factor and the interaction between test stimuli and percentage of reinforcement were not statistically significant. (c) Extinction occurred rapidly so that the gradient failed to reach statistical significance beyond the 1st test trial. These results, when combined with similar findings from other studies, make the usefulness of extinction or "cycling" test trials seem of doubtful value in eyelid conditioning with present techniques.

Primary stimulus generalization along a continuum of intensity has been demonstrated in classical conditioning with the salivary response (Razran, 1949) and the galvanic skin response (Hovland, 1937). In 54 Russian experiments summarized by Razran, gradients obtained for postacquisition tests showed monotonically decreasing response strengths for stimuli of decreasing intensity and monotonic increases in level of responding when tests were conducted for stimuli with intensities greater than the CS. Comparable results were obtained by Hovland. Eyelid conditioning studies in which regular gradients of generalization were obtained in either quality or intensity dimensions have been generally absent from the literature. A survey of the eyelid conditioning experiments involving generalization reveals two studies in which nonsystematic responses to test stimuli were obtained (Grant & Schneider, 1948; Hake, 1948), one study in which a tendency toward systematic decrements was observed along an intensity dimension (Taylor, 1954), and one investigation involving a differential effect of drive level on responses through extinction to one auditory test stimulus differing in frequency from the CS (Porter, 1962).

Grant and Schneider assigned *Ss* to 1 of 16 cells in a 4 × 4 factorial design in which acquisition trials were to one of four light intensities and extinction was carried out to the same or a different intensity. The basic purpose of the study was the investigation of the effects of CS intensity upon learn-

---

[1] This research is based on a problem submitted to the University of Massachusetts in partial fulfillment of the requirements for the MS degree by the senior author under the direction of the junior author. A portion of this experiment was reported at the Eastern Psychological Association, New York, April 1963.

The authors are indebted to A. E. Goss for his generous help and guidance throughout the course of the investigation.

ing and performance with generalization as one variable. Both the Hake and Taylor studies used tones as the CS with the former employing a differential conditioning procedure during acquisition and a cycling procedure during test trials for primary stimulus quality generalization. The Taylor investigation involved assignment of *Ss* during test trials to various intensities of the tone used as the CS and obtained a nonsystematic gradient for the first postacquisition test trial but more systematic decrements when the first and second test trials were combined.

This investigation is concerned with demonstrating a gradient of generalization along a continuum of intensity in eyelid conditioning. The CS was a light of relatively high intensity and testing was carried out with the CS and three stimuli of proportionately less intensity. Another variable investigated was the effect of partial vs. 100% reinforcement upon generalization. Following acquisition "pure" generalization, i.e., without the confounding effects of extinction, was studied on the first test trial and the effects of experimental extinction upon the gradient were observed through further nonreinforced trials.

METHOD

Apparatus.—The *Ss* were located in a dimly illuminated 6 ft. 6 in. × 7 ft. 5 in. room. This room was adjacent to *E*'s room and contained a fan to make any noise emanating from equipment in *E*'s room inaudible. The *Ss* were seated in a partially reclined barber's chair with attached headrest and faced the stimulus unit, a flat black panel with dimensions of 3 × 3 ft. A circular milk-glass disk with a diameter of 6 cm. was located in the center of this panel which was a distance of 60 in. from *S*'s eye. The duration of the stimuli and the interval between them were controlled by Hunter interval timers with the intertrial interval manually controlled by *E*. Recording of *S*'s eyelid movements was made by means of a microtorque potentiometer, Hunter amplifier, and Epsco penmotor. The penmotor was mounted on a Brush paper-puller which pulled paper through the recording system at a rate of 120 mm/sec.

Design.—The *Ss* were assigned to eight cells in a 2 × 4 factorial design in which two percentages of reinforcement were orthogonal to four test stimuli.

Procedure.—Following necessary adjustments to the apparatus instructions designed to leave *Ss* with a passive set were read aloud to *S*. Ninety acquisition trials were then given. The CS consisted of an increase in brightness of the milk-glass disk from .032 to 3.20 apparent ft-c. The UCS was an air puff of 2.0 psi released by a solenoid valve and delivered to the right eye by means of a nozzle attached to a headset worn by *S*. For all reinforced trials the CS was of 550 msec. duration. The UCS was of 50 msec. duration and followed the onset of the CS by 500 msec. on reinforced trials, both stimuli terminating simultaneously. This procedure remained constant throughout all 90 acquisition trials for those *Ss* receiving 100% reinforcement. On half of the trials for the 50% reinforcement group the CS was of 2,550 msec. duration with the UCS following the onset of the CS by 2,500 msec., an interval which has been found not to produce stable conditioning while maintaining drive level (McAllister, 1953). Reinforced and nonreinforced trials were alternated randomly with the restrictions of equal occurrence in successive blocks of 10 trials and no more than three consecutive presentations of each. The intertrial interval was randomly varied from 10 to 20 sec. during both the acquisition and extinction trials.

Twenty test-extinction trials with the nonreinforcing delayed stimulus conditions described earlier were given following acquisition. Test stimuli intensities were 3.20, 1.364, .4267, and .07459 apparent ft-c from a background light of .032 apparent ft-c. These values obtained by means of a Variac transformer and Wratten neutral density filters represent steps of equal psychological magnitude between the background light and the most intense stimulus as computed from the "psychophysical power law" proposed by Stevens (1962). Measurements were made from the source with a MacBeth illuminometer. The *Ss* were randomly placed in either the 100 or 50% reinforcement conditions prior to acquisition trials. Assignment to the various test stimuli within each percentage of reinforcement was based on the number of CRs given during the last 10 acquisition trials and was designed to insure that the groups were matched in response strength for the test trials.

Subjects.—The *S*s were 80 male students from introductory psychology classes at the University of Massachusetts.

Definition of response.—A conditioned response was defined as a deflection of 1 mm. of the recording pen occurring between 200 and 525 msec. after the onset of the CS. The critical area for CRs was extended 25 msec. beyond the onset of the UCS as the minimum reaction time to the UCS was determined as 35 msec. for 10 *S*s who were each given 10 trials with the UCS alone. Fourteen additional *S*s were eliminated. Three of these were judged as voluntary responders having met the criterion of greater than 50% of their responses occurring between 200 and 300 msec. following the onset of the CS. Two *S*s from the 100% group and 9 from the 50% group were eliminated from the data as having failed to meet arbitrary criteria for response strength. In order to insure a high level of response strength for the generalization data these criteria of 50% CRs and 30% CRs during the last 10 acquisition trials were employed for the 100 and 50% reinforcement groups, respectively.

RESULTS

Acquisition.—All groups displayed increases in the number of responses given in each block of five trials as acquisition progressed with the 100% reinforcement group being superior to the 50% group both in the final level reached in the last five trials and in rate of approach to asymptote. Within each percentage of reinforcement the groups tested with different stimuli during extinction responded at comparable rates.

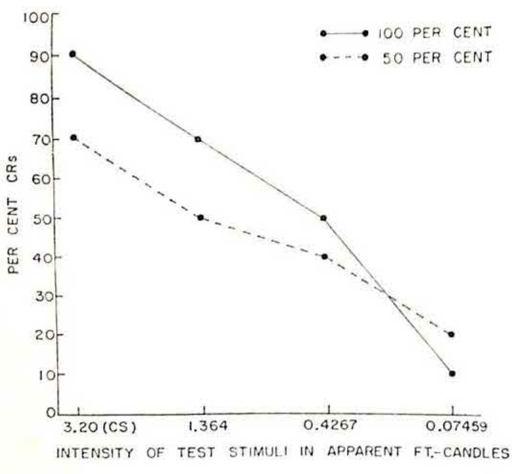


FIG. 1. Gradients of generalization for first test trial for 100 and 50% groups.

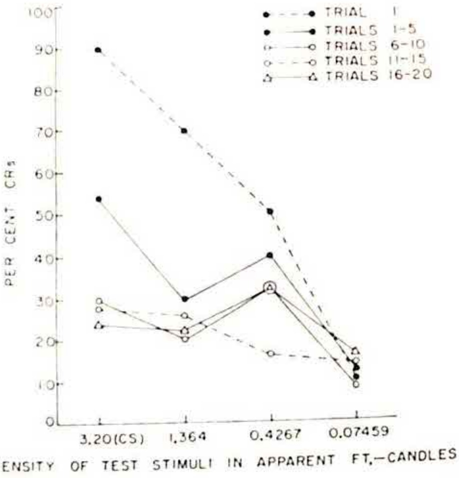


FIG. 2. Generalization gradient for 100% group during extinction.

An analysis of variance of these acquisition data showed a significant Trials effect at less than the .001 level. The significant Percentage of Reinforcement factor ($p < .01$) and the Percentage × Trials interaction ($p < .001$) show that the 100% group differed from the 50% group in level of responding and in approach to asymptote. The nonsignificant Stimuli factor ($p > .20$) and Percentage × Stimuli interaction ($p > .20$) substantiate the observation that within each percentage of reinforcement the groups did not differ significantly in response strength at the start of generalization and extinction.

Generalization.—Figure 1 shows that the gradients of generalization on the first test trial for both 100% and 50% groups are systematic in their reflection of the decremental effects of lessened intensity from the CS. The gradient for the 100% group appears to have a steeper slope than that of the 50% group.

Data were analyzed by means of a Sutcliffe chi square test (Sutcliffe, 1957), a method for utilizing a chi square technique with more than one variable. The Stimuli effect was significant, $\chi^2 (3) = 18.00$, $p < .01$, reflecting generalization factors.

Neither Percentage of Reinforcement nor the interaction of Percentage of Reinforcement with Stimuli approached significance. Separate chi square analyses reveal a significant Test Stimuli effect for the 100% group, $\chi^2 (3) = 14.14$, $p < .01$, but nonsignificance for the 50% group, $\chi^2 (3) = 5.25$, $p > .10$.

For the second trial the decremental effects of distance from the CS have largely disappeared due to the rapid extinction of responses to the CS and the stimulus nearest it. The Sutcliffe analysis reveals no significant effects. A chi square analysis performed with the 100% group also failed to indicate any significant Test Stimuli effect for this trial, $\chi^2 (3) = 2.20$, $p > .50$.

Extinction.—Figures 2 and 3 show that the systematic gradient rapidly disappears with nonreinforced trials, showing a nearly horizontal shape for both 50 and 100% groups during the first block of five trials. This is due primarily to the continued extinction of the CS and the 1.364 groups whereas the group receiving Test Stimulus .4267 actually gained some response strength as trials progressed. The .07459 group remained at its original low level throughout extinction. Extinction occurred rapidly as the combined groups responded at a 50% level on the first test trial and at 34% for Trials 1 through 5. By the second block of five extinction trials, the percentage of CRs had dropped to 26%, below the level for the first block of five acquisition trials. An analysis of variance reveals a significant Trials effect during extinction ($p < .01$) as shown in Table 1. The Stimuli × Trials interaction ($p < .01$) reflects the previously noted differential effect of extinction upon the various test stimuli. As a consequence, the significant Stimuli effect ($p < .05$) cannot be interpreted as reflecting the persistence of a systematic gradient throughout extinction.

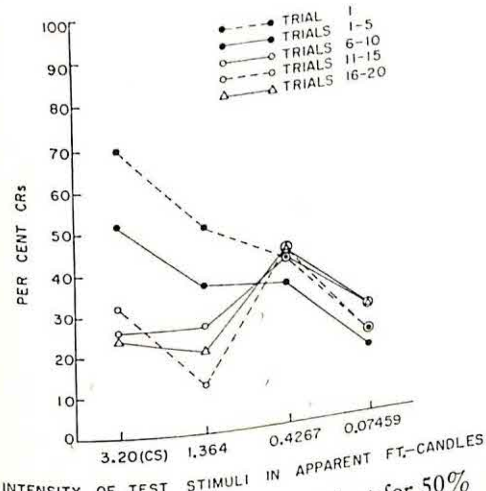


FIG. 3. Generalization gradient for 50% group during extinction.

TABLE 1

ANALYSIS OF VARIANCE ON CRs DURING EXTINCTION

Source	*df*	*MS*	*F*
Percentage of reinforcement (P)	1	3.828	<1
Stimuli (S)	3	13.928	2.86*
P × S	3	1.753	<1
Error (b)	72	4.877	
Trials (T)	3	4.328	4.82**
P × T	3	.396	<1
S × T	9	2.436	2.71**
P × S × T	9	1.167	1.30
Error (w)	216	.899	

* $p < .05$.
** $p < .01$.

DISCUSSION

All test stimuli groups within each reinforcement condition entered test trials with nearly equivalent response strength, the differences between groups being statistically nonsignificant. During the last five acquisition trials, the groups receiving 100% reinforcement gave 80.5% CRs while the partial reinforcement groups gave 65.5% responses. Though the latter groups were elevated by the elimination from the data of nine *S*s responding at a low level, the percentage of reinforcement variable reached statistical significance. In interpretation of the generalization gradients, however, the lack of differential effect of reinforcement schedule is restricted by the criterion procedure.

The monotonically decreasing gradient of generalization found in the present investigation is in contrast to the results of other generalization studies in eyelid conditioning (Grant & Schneider, 1948; Hake, 1948; Taylor, 1954). Grant and Schneider found no systematic decremental effects on either frequency of response or amplitude measures when testing along an intensity dimension during extinction. Hake obtained nonsystematic results when tests utilizing a cycling technique were conducted with tones of various frequencies following a differential conditioning procedure. The results of both of these studies include extinction factors as well as generalization. The gradients of the present study lost their regularity rapidly in extinction, departing from a monotonically decreasing shape during the first block of five extinction trials and becoming horizontal during the second trial block. These findings and the usual rapid extinction of the eyelid CR (Gormezano & Moore, 1962; Taylor, 1954) indicate that test trial data beyond the first postacquisition trial reflect extinction factors to a degree which may obscure the effects of other variables.

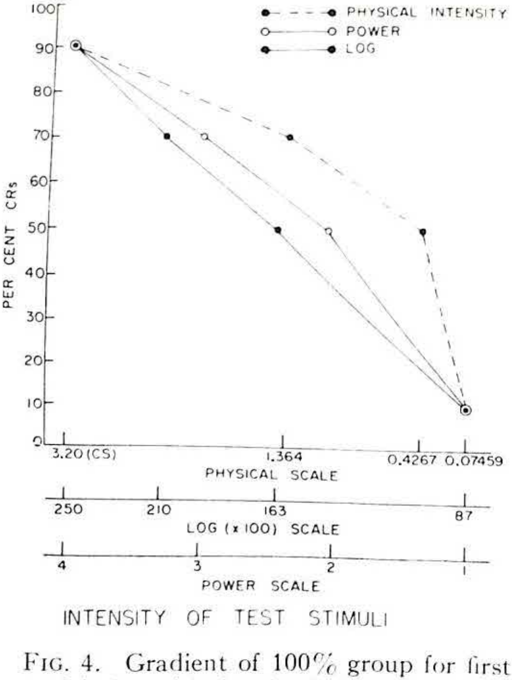


FIG. 4. Gradient of 100% group for first test trial plotted in log (× 100), power function, and physical intensity scales.

Taylor (1954) utilized auditory stimuli and tested upward along the intensity dimension. Previous classical conditioning studies in which testing has been carried out in this direction have not found any decrements in response strength to those more intense stimuli (Hovland, 1937; Razran, 1949). The point of interest might be that Taylor observed decremental effects where others had not, rather than that the gradient for the first test trial was not monotonic. Taylor comments that the GSR used by Hovland is more sensitive to stimulus intensity than is the eyelid response and notes a high percentage of short latency responses occurring to the more intense of her test stimuli. It is possible that stimulus intensity has a differential effect upon these two responses as Walker (1960), in studying the effects of stimulus intensity on acquisition and extinction in eyelid conditioning, also observes a shortening of response latencies at the higher stimulus intensities.

Previous investigations have used either a log scale or a summation of jnd's to estimate psychological magnitudes, whereas in the present study the power function proposed by Stevens was employed. White (1962) has reviewed scaling methods with regard to the asymmetry of gradients along an intensity dimension replotting some of the data in the literature with a power function baseline. As a result of this type of rescaling, the gradients obtained by Hovland (1937) do not show as pronounced a tendency toward being linearly related to stimulus intensity. Figure 4 displays the gradient obtained in the present study for the 100% groups on the first test trial using a power function baseline and the values as they would fall on log and physical intensity scales. The gradient is of almost perfect linear shape if the log scale is used but appears concave downward plotted over the power function and physical intensity baselines. Since the shape of the gradient varies considerably according to the scaling procedure employed it might be parsimonious to disregard scaling procedures in generalization and study the parameters relating to some physical dimension as has been previously accomplished in eyelid conditioning with UCS intensity (Ross & Spence, 1960) and CS-UCS interval (McAllister, 1953; Kimble, Mann, & Dufort, 1955).

REFERENCES

Gormezano, I., & Moore, J. W. Effects of instructional set and UCS intensity on the latency, percentage, and form of the eyelid response. *J. exp. Psychol.*, 1962, **63**, 487–494.

Grant, D. A., & Schneider, D. E. Intensity of the conditioned stimuli and strength of conditioning: I. The conditioned eyelid response to lights. *J. exp. Psychol.*, 1948, **38**, 690–697.

Hake, H. W. Generalization and discrimination in the conditioned eyelid response to sound stimuli. Unpublished master's thesis, University of Wisconsin, 1948.

Hovland, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–291.

Kimble, G. A., Mann, L. I., & Dufort, R. H. Classical and instrumental eyelid conditioning. *J. exp. Psychol.*, 1955, **49**, 407–417.

McAllister, W. R. Eyelid conditioning as a function of the CS-US interval. *J. exp. Psychol.*, 1953, **45**, 417–422.

Porter, J. J. Stimulus generalization as a function of UCS intensity in eyelid conditioning. *J. exp. Psychol.*, 1962, **64**, 311–313.

Razran, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.

Ross, L. E., & Spence, K. W. Eyelid conditioning performance under partial reinforcement as a function of UCS intensity. *J. exp. Psychol.*, 1960, **59**, 379–382.

Stevens, S. S. The surprising simplicity of sensory metrics. *Amer. Psychologist*, 1962, **17**, 29–39.

Sutcliffe, J. P. A general method of analysis of frequency data for multiple classification designs. *Psychol. Bull.*, 1957, **54**, 134–137.

Taylor, E. C. N. Generalization of the conditioned eyelid response to an auditory stimulus varying in intensity. Unpublished doctoral dissertation, State University of Iowa, 1954.

Walker, E. G. Eyelid conditioning as a function of intensity of conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1960, **59**, 303–311.

White, S. N. A note on intensity generalization and prothetic scaling. *Psychol. Rev.*, 1962, **69**, 149–155.

(Received March 15, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 168–172

EFFECT OF REDUNDANCY AND DURATION ON ABSOLUTE JUDGMENTS OF VISUAL STIMULI[1]

W. R. GARNER AND C. DOUGLAS CREELMAN[2]

Johns Hopkins University

Absolute judgments of stimuli which varied in size, in hue, or in size and hue combined in a correlated fashion, were made at 2 different durations. The results showed considerable gain in discrimination with combined size-hue stimuli, but no effect of duration at all. It is suggested that such tasks are of judgmental discrimination rather than perceptual discrimination, and that while absolute judgments can be limited by perceptual factors, other factors are ordinarily more important.

Very few experiments have shown that perceptual discrimination of a redundant set of stimuli is better than discrimination of a nonredundant set of the same size. In other words, the addition of one or more correlated dimensions to a set of stimuli does not ordinarily improve discrimination of the stimuli. An important exception to this generalization is the finding by Eriksen and Hake (1955b) that combinations of size, hue, and brightness in an absolute discrimination task gave better discrimination, as measured by information transmission, than any of the dimensions judged singly. Because this result is so at variance with other results on perceptual discrimination, a better understanding of the nature of this task would be worthwhile.

The present experiment repeats part of the Eriksen and Hake experiment, except that stimulus duration was introduced as an additional experimental variable, in order to determine whether the earlier result was due to the lack of controlled duration.

Our interest in duration as an experimental variable in this type of problem stems from two causes. First, Garner (1962, p. 209) has suggested that the gain in discrimination with the addition of a redundant variable might not occur with short durations of stimulus exposure. In the original Eriksen and Hake experiment, the stimulus was exposed as long as necessary for *S* to make his judgment, and there is a possibility that the gain due to redundancy occurred at the cost of increased time.

Second, Garner and Lee (1962) have analyzed in detail the factors which can affect the utility of redundancy in perceptual discrimination tasks. One critical factor in the role of redundancy is the *perceptual independence* of the psychological counterparts of the stimulus dimensions which are made redundant, or correlated, with each other. Only if the perceptual counterparts are independent can there be a gain due to the addition of a correlated dimension. But more important for our immediate purpose,

---

[1] This research was done under Contract Nonr-248(55) between the Office of Naval Research and the Johns Hopkins University. This is Report No. 24 under that contract. Reproduction in whole or in part is permitted for any purpose of the United States Government. The authors wish to express their appreciation to William Ertag for assistance in the collection and analysis of the data.

[2] National Institute of Mental Health Postdoctoral Fellow.

these authors also pointed out that stimulus duration itself can be considered as a kind of redundancy, since an increase in duration is essentially the repetition of the same stimulus. In other words, two or more contiguous time periods can be considered to be providing correlated stimulus variables. Thus redundancy can be produced either by increasing duration or by adding a correlated dimension at a fixed duration, and a better understanding of the problem can be obtained by studying the joint effects of duration and number of dimensions in discrimination tasks.

METHOD

Stimulus conditions.—The *S*s made absolute judgments of three different sets of 20 squares, presented visually at one of two durations. One set of squares varied in size alone, another in hue alone, and a third set varied in both size and hue, with the two dimensions correlated.

Stimuli for size judgments were squares ranging from $\frac{1}{4}$ to 5 in. on a side, in $\frac{1}{4}$-in. steps, cut from black Munsell paper. These stimuli were twice as large as those used by Eriksen and Hake because we used a much greater viewing distance.

Stimuli for the hue series were $1\frac{3}{4}$-in. squares of Munsell paper. The hues were 5R, 10R, 5YR, 10YR, 5Y, 10Y, 5GY, 10GY, 5G, 10G, 5BG, 10BG, 5B, 10B, 5PB, 10PB, 5P, 10P, 5RP, 10RP; and in each case the Munsell value was 5, and the chroma was 6. These are the same hues used by Eriksen and Hake.

The correlated stimulus set varied both in size and in hue, with the two dimensions related in a one-to-one manner. Again, this series was the same as that used by Eriksen and Hake, except for the larger sizes.

The numbers 1 through 20 were used as responses for each set of squares. In the case of the size series the smallest square was called 1, and the largest 20. In the hue series, the numbers were assigned in order through the spectrum, with the response 1 associated with 5R, and so on. In the correlated, redundant stimulus series, the smallest square was 5R, the next smallest was 10R, and so on, with the same response numbers as those used with the single-dimensional stimuli.

Two different durations of stimulus exposure were used— $\frac{1}{25}$ and $\frac{1}{10}$ sec. Each stimulus series was presented at both durations, giving a total of six conditions.

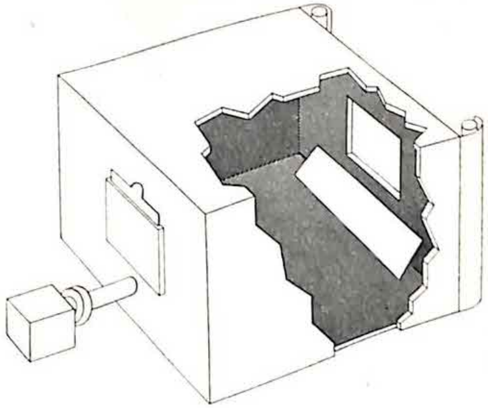


FIG. 1. A schematic drawing of the reduction box and projector apparatus used in the experiment.

Apparatus.—In order to control the duration of the stimuli, fix the position, and introduce small color distortion, we needed a more elaborate apparatus than that used by Eriksen and Hake. Each stimulus was mounted on a separate white card 10 in. square, which could be inserted in the back of a reduction screen box. Figure 1 is a schematic drawing of this part of the apparatus. The box was 22 in. deep, and 28 in. square as seen from the front (at the right of the figure). The aperture in the front of the box was 10 in. square, and the inside of the box was lined with black, light-absorbent material.

The stimulus card was illuminated for the required duration by a light projected through the box from behind, aimed to reflect off a diffusing white board located inside the box, below the viewing aperture. The duration was controlled by a shutter on the light projector.

The front surface of the reduction box was painted flat white and was illuminated from two sides by masked fluorescent bulbs mounted in reflectors. The bulbs themselves could not be seen by *S*, but they gave a reasonably uniform light across the white surface. The brightness of the white surface was approximately 3 apparent ft-c, and the brightness of the illuminated stimulus card at the back of the reduction box was the same. The lighted front surface was used to help maintain visual adaptation for *S* and to minimize the possibility of afterimages.

A portion of the apparatus not shown in

Fig. 1 was a 10-ft. tunnel, which extended to the right of the figure. The tunnel was itself lined with black felt to prevent light from reflecting off the walls back onto the stimulus cards. The *S* was seated at the end of the tunnel, and viewed the face of the reduction box through a binocular viewing aperture. What *S* saw was a white field with a black 10-in. square in the center. He saw the stimulus on its white card only during the brief exposure period, when the shutter of the projector was activated. The only purpose of the reduction box and tunnel was to ensure that the stimulus cards could be changed between trials, but be seen by *S* only during the proper exposure period.

Subjects and procedure.—Six undergraduate students served as *Ss*, and each *S* went through all experimental conditions. The six conditions were assigned to the six *Ss* in a Latin square counterbalanced order, and each *S* continued with one condition until all judgments had been made.

The stimuli were presented in random order approximately one every 10 sec. A warning bell rang 3 sec. before each stimulus presentation. During the time between stimulus presentations, *E* removed the stimulus card from its holder on the reduction box and inserted another, selected from a nearby rack. After each stimulus, *S* wrote his response on a sheet of paper and also called it out to *E*, who then told him what the correct response was. Each *S* made a total of 36 judgments of each stimulus for each condition, but the first 6 judgments were considered practice, and analysis was done only for the last 30 judgments. The *S* served no more than an hour at a time, and was given a brief warm-up period at the beginning of each hour.

Results

Information transmission between stimuli and responses was computed separately for each *S* and each condition, and the averages of these values for each of the six conditions are shown in Table 1. Analysis of variance showed that the stimulus set (number and type of dimension) was a highly significant source of variance, but that neither the main effect of duration nor the interaction of duration and stimulus set was a significant source of variance.

TABLE 1

INFORMATION TRANSMISSIONS FOR THE THREE STIMULUS SETS AND THE TWO DURATIONS

Duration (Sec.)	Stimulus Set			
	Hue	Size	Hue-Size	*M*
1/10	2.39	2.69	3.32	2.80
1/25	2.30	2.68	3.27	2.75
M	2.35	2.68	3.30	2.78

Note.—Each information transmission is the mean, in bits, for six *Ss*.

Our values of information transmission for size judgments are very close to those obtained by Eriksen and Hake. Our values for the hue series are somewhat lower, but this difference can be accounted for by our use of a yellow illuminant (a projection bulb rather than a Munsell lamp) and by the fact that some of our *Ss* were slightly color weak.

The critical result, however, is that information transmission was substantially higher when both stimulus dimensions were used together than when either was used alone, thus confirming the earlier result of Eriksen and Hake even with short durations. An equally critical result is that there was no difference between the two durations.

Discussion

Duration had no effect on accuracy of absolute judgment in this experiment, but adding a correlated stimulus dimension did. Each of these facts is of importance for understanding the nature of the task.

First, the fact that adding a correlated dimension improved judgmental accuracy even when short durations were used means that the original Eriksen and Hake result was not an artifact due to their allowing *S* an unlimited time for judgment. The gain due to redundancy does not, then, depend on *S*'s using additional time, as originally suggested by Garner (1962). Furthermore, this gain due to redundancy indicates, according to the Garner and Lee (1962) analysis, that the two stimulus dimensions of size and hue have perceptual counterparts which are independent. The Garner and Lee analysis shows that the maximum possible gain due to the addition of a redundant variable is .5 bit, and in the present experiment this value is exceeded, although not significantly so. Thus as much gain was obtained as is theoretically possible.

In discussing this problem, Garner and Lee used the term *perceptual independence*, meaning that errors with regard to the two dimensions are uncorrelated. In other words, whatever factors lead to less than perfect discrimination with one dimension are not the same, or are not correlated with, those that limit discrimination with the other dimension. These factors need not, however, be perceptual factors in the sense that there are limiting conditions at the receptor input end of the total psychological process. The limiting factors can just as well be judgmental in nature, having to do with such things as *S*'s inability to keep stimuli and responses aligned, or to maintain a stable frame of reference.

Second, the fact that no effect of duration was obtained in this experiment suggests that the limiting factor on performance is not from difficulty in discrimination, but rather is judgmental. The Garner and Lee experiment indicated that duration changes of the order used in this experiment should lead to a difference in information transmission of approximately .6 bit. We obtained a nonsignificant difference of .05 bit, and even though it is in the expected direction, it is clearly far from what one should expect if the task is truly perceptual in its limitations.

It is interesting that there have been, to our knowledge, no experiments directly concerned with the effect of short duration on absolute judgments. We assumed that duration would have a strong effect, as it has on practically all discrimination tasks. And yet in retrospect, we should not have been surprised at the lack of such an effect. A number of studies have shown that factors which should affect discrimination have little effect on absolute judgment. Pollack (1952), for example, showed that restriction of the range of frequencies in absolute judgments of pitch had little effect on information transmission; Garner (1952) showed the same for judgments of loudness. Eriksen and Hake (1955a) and Alluisi and Sidorsky (1958) have shown the same lack of effect for judgments of visual size.

Yet such restrictions of range should have a major effect if the critical factor in the task were perceptual discrimination between stimuli. Furthermore, at least in the case of loudness judgments, the effect of duration should be much the same as the effect of restriction of range, since to reduce duration is essentially to reduce the loudnesses of all stimuli proportionately, i.e., to reduce the absolute range of loudnesses.

Perceptual discrimination or judgmental discrimination?—The fact that absolute judgment is so little affected by factors which do affect perceptual discrimination leads us to believe that absolute judgment is not primarily or necessarily a task of perceptual discrimination. In the typical absolute judgment experiment, stimuli arranged along a single continuum (or even a correlated pair of continua) are used, but the adjacent stimuli on the continuum would not be confused if they were presented for pair discrimination. Yet when a large set of such stimuli is used, adjacency discrimination is far from perfect. Furthermore, total discrimination as measured by information transmission is unaffected by ordinary perceptual factors.

Absolute judgment tasks are operationally like perceptual discrimination tasks, but this operational similarity does not mean that they are in fact the same. The absolute judgment task is more concerned with judgmental discrimination than with perceptual discrimination, and different laws will hold for these two different discrimination tasks. For perceptual discrimination, factors which affect the receipt of information at the receptor will be critical. But these same factors will not be critical for judgmental discrimination. In judgment, such things as the nature of the response continuum, the compatability of the response to the stimulus, and problems of decision making will be critical. Since, however, the absolute judgment task does require an initial perceptual input, then it will be limited by factors which affect perceptual discrimination. But the judgment task itself will impose limits on discrimination well beyond those imposed by perceptual factors, and these will ordinarily be the overriding factors.

REFERENCES

Alluisi, E. A., & Sidorsky, R. C. The empirical validity of equal discriminability scaling. *J. exp. Psychol.*, 1958, **55**, 86–95.

Eriksen, C. W., & Hake, H. W. Absolute judgments as a function of stimulus range and number of stimulus and response categories. *J. exp. Psychol.*, 1955, **49**, 323–332. (a)

Eriksen, C. W., & Hake, H. W. Multidimensional stimulus differences and accuracy of discrimination. *J. exp. Psychol.*, 1955, **50**, 153–160. (b)

Garner, W. R. An equal discriminability scale for loudness judgments. *J. exp. Psychol.*, 1952, **43**, 232–238.

Garner, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

Garner, W. R., & Lee, W. An analysis of redundancy in perceptual discrimination. *Percept. mot. Skills*, 1962, **15**, 367–388.

Pollack, I. The information of elementary auditory displays. *J. Acoust. Soc. Amer.*, 1952, **24**, 745–749.

(Received March 16, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 173–182

COMPARISON OF MEANINGFULNESS AND PRONUNCIABILITY AS GROUPING PRINCIPLES IN THE PERCEPTION AND RETENTION OF VERBAL MATERIAL[1]

ELEANOR J. GIBSON, CAROL H. BISHOP, WILLIAM SCHIFF, AND JESSE SMITH

Cornell University

Thresholds of visual perception and 2 measures of retention were obtained for trigrams varying in pronunciability and meaningfulness (semantic reference of the kind found in well-known initials), and for control items. The 3 types of trigrams contained the same letters rearranged anagram-wise. Perceptual thresholds were lowest for pronounceable items, and next lowest for meaningful ones. On the other hand retention, measured by both recognition and free recall, was best for the meaningful items and second best for pronounceable ones. Pronunciability was inferred to be the better grouping principle for reading or coding to speech units. Meaningfulness was inferred to have facilitated retention more than pronunciability by providing a category for grouping the initial items, thus aiding retrieval.

Hardly anyone would question the proposition that the efficient reader or thinker deals with larger units of material, whatever they may be, than the beginning or retarded practitioner. It was not necessary for psychologists to point out that speedy readers do not dwell on every letter, but take in words, phrases, and conceivably even sentences "at a glance." Unlike some common sense knowledge, however, this maxim has received convincing corroboration from well-conducted experiments. Perhaps the most famous of these is Bryan and Harter's (1899) which demonstrated that progress beyond a certain point in the reception of Morse code depended on the development of higher-order units. The importance of higher units in the perception of verbal material was early demonstrated by Cattell (1885) and by Erdmann and Dodge (1898). The number of letters which can be perceived in a tachistoscopic exposure increases as the letters form a word, and again as the words form sentences.

For several decades psychologists accepted these facts with satisfaction and with little further curiosity. But in recent years, new interest has centered on the problem of how the higher units are formed. At least three lines of investigation have led to this revival. One is the advent of information theory and the realization that information processing is affected by what Miller (1956) called "chunking." Another is the emphasis of rote learning theorists on "response integration" (Mandler, 1954, 1962; Underwood & Schulz, 1960); and third, experiments on the reading process considered as perceptual learning (Gibson, Osser, & Pick, 1963; Gibson, Pick, Osser, & Hammond, 1962).

The general questions to be answered are several. How are higher-orders units formed? In fact, are they formed by the learner, or are they rather discovered in the structure of the stimulus array? What principles of grouping, in either case, favor the development of higher units, and are these principles the same, whatever the task to be performed with the units?

---

[1] This work was supported in part by a grant from the United States Office of Education, Department of Health, Education, and Welfare.

It was generally taken for granted in the early work on tachistoscopic perception of word vs. nonword sequences of letters that words were read faster because they were "meaningful"; meaning presumably conferred unity on the letter group. But recent work (Gibson et al., 1962; Underwood & Schulz, 1960) makes plausible the hypothesis that pronunciability—the correlation of a spelling pattern with an invariant pronunciation—is an alternative and important grouping principle for letter sequences, a cohesive property making for unit formation.

Is it possible to separate these two properties in order to compare their effectiveness? The difficulty of defining meaningfulness (or at least agreeing on a definition satisfactory to more than a few people) is at first discouraging. Furthermore, real words, however meaningful by any criterion, are de facto pronounceable. And nonsense syllables, when rated for meaningfulness and for pronunciability, appear to have these properties correlated. This fact is not surprising, for pronunciability is itself a kind of meaning. A pronounceable letter group obeys rules of English grapheme-phoneme correspondence and therefore has a structure which is familiar and related to other verbal structures. But it may nevertheless lack semantic reference and thus be low in that aspect of meaning.

A method of separating meaning in the sense of semantic reference from pronunciability was suggested to us by the experiments of Underwood and Schulz (1960). They found that pronunciability of trigrams was an excellent predictor of learning, but there were some notable exceptions. These were letter sequences which had either extremely high generated value (i.e., xyz), or very consistent and familiar semantic reference (i.e., initials such as RFD). These two indexes of meaning are quite different from the one implied by Noble's (1963) *m* (an index of meaningfulness given "by the average number of associations made by a representative sample of subjects during a standard time interval [p. 84]"), but they can nevertheless be defended as equally relevant. Underwood and Schulz (1960) considered generated value of the kind found in xyz a useful measure of meaningfulness. More specifically, they said "We are interpreting both pronunciability and the generated values to be measures of response integration [p. 246]." As for consistent semantic reference, it may be that a letter group is no less meaningful when it has one and only one association which everyone agrees on than when it has a numerous and varied set of them.

Familiar initials satisfy both these criteria of meaningfulness; they have high generated frequency of the third letter, given the first two, and they have consistent semantic reference. Furthermore, a number of them defy pronunciation as a unit, so that the letter sequence can be considered high in one property and low in the other. Underwood and Schulz said, in comparing generated values with pronunciability, "Thus if we capitalize on alphabetical-sequence habits, initial habits, and so on, which often result in trigrams difficult to pronounce, we could quite easily reverse the relative predictive capacity shown for the two measures in the present experiment [p. 258]."

Whether there are differences in the effectiveness of verbal units "packaged" primarily by meaningfulness as just defined, or by pronunciability is the subject of this investigation. Their effectiveness as grouping principles is compared by determining perceptual thresholds for items high in either pronunciability or in meaningfulness, but low in the other, and for control items low in both. Their effectiveness is also compared for two measures of retention.

METHOD

Material.—Three types of trigrams were selected for comparison: trigrams whose unitariness was constituted primarily by their being pronounceable as a monosyllabic unit, according to the rules of English spelling and pronunciation; trigrams whose unitariness was constituted primarily by meaningfulness defined by semantic reference and generated value; and control trigrams low in both pronunciability and meaningfulness.

Since we considered it important to control the letter shapes, which are themselves a source of error in measuring perceptual thresholds, the trigrams were chosen so that each control trigram, rearranged anagramwise, yielded a pronounceable counterpart and a meaningful counterpart. For example, a control trigram MBI had as its pronounceable counterpart MIB and as its meaningful counterpart IBM.

First a search was made for trigrams yielding a triad of the kind described. Next, all these items were checked for generated frequency of the third letter in the tables in Underwood and Schulz (1960, pp. 377 ff.), since it was desired that meaningful items be high in this respect and the others low. When there was a choice between two control or two pronounceable trigrams (e.g., MIB or BIM), the one with lower generated frequency of the third letter was chosen.

After these steps, we were left with 15 sets of trigrams to be rated for meaningfulness by a population similar to that finally to be used as *S*s. Subordinate lists of 15 trigrams were constructed, each containing only one member of a triad such as IBM, MIB, and MBI. Two versions of each of the three resulting lists were made with different random orders. Five real words (the same ones) were inserted in each of the lists as well. Because the first tryout indicated that some of our would-be meaningful items might turn out to be low for the population tested and would have to be discarded, four more such items were added to all the lists to be rated. Each list finally included nine supposedly meaningful trigrams, five pronounceable trigrams, five control trigrams, and five words.

TABLE 1

PERCENTAGE OF *S*S RATING TRIGRAMS AS VERY MEANINGFUL OR SOMEWHAT MEANINGFUL

Control		Pronounceable		Meaningful	
Trigram	%	Trigram	%	Trigram	%
OKR	9	KOR	5	RKO	74
AVT	9	TAV	6	TVA	88
MBI	3	MIB	12	IBM	100
IFB	7	BIF	11	FBI	100
ACB	12	BAC	16	ABC	95
ATW	0	TAW	3	TWA	93
RLI	0	RIL	8	ILR	95
APW	3	WAP	24	WPA	71
OTK	4	TOK	3	TKO	91
YNU	7	YUN	6	NYU	94
ICF	—	CIF	—	IFC	95
LFA	—	FAL	—	AFL	93

The six subordinate lists were distributed to different freshman English classes. Each item was rated for meaningfulness on a 4-point scale (very meaningful, somewhat meaningful, slightly meaningful, not at all meaningful). In addition, the rater was asked to indicate the meaning of each item, if it had any, in order to locate any idiosyncratic items. Each trigram was rated by at least 30 to 40 people.

Table 1 shows the 12 sets of trigrams finally chosen, with the percentage of *S*s who rated the trigrams very meaningful or somewhat meaningful. Of the 19 items intended to be meaningful, 7 had to be discarded, either because they were not rated high enough by a majority of *S*s, or because one of the two counterparts was rated high. Two of those retained were ones inserted after the first tryout and did not, therefore, have their counterparts rated. The meanings indicated by *S*s were examined for idiosyncrasy. Items retained as meaningful had consistent reference, high generated value, and a high meaningfulness rating.

Ratings of pronunciability were not obtained. Instead, the trigrams were presented to a group of judges (members of a graduate seminar in verbal learning). There was unanimous agreement that the items in the pronounceable list were pronounceable as a monosyllabic unit according to English usage, and were also easier to pronounce, by a native English speaker, than their two counterparts.

Lists to be used for testing perceptual thresholds and retention were "mixed," that is, made up with four items of each of the three groups. A list contained only one member of a triad. All items appeared in a list an equal number of times and were distributed so as to randomize list membership and so as to counterbalance for possible position effects within a list. Order within the list was random, but with the stipulation that items of the same type be separated by two other items. Lists of items presented for threshold determination included four real words interspersed among the trigrams. Ratings of meaningfulness for these words (SUM, HOG, RAT, JOY) were lower on the average than for the meaningful trigrams, suggesting that a successful effort had been made to achieve high meaningfulness for these trigrams.

Method for determining perceptual thresholds.—Thresholds were measured individually for each *S* by means of varying the brightness contrast of the exposure field in relation to the pre- and postexposure fields. The time of exposure was maintained constant at .25 sec. Brightness of illumination of the pre- and postexposure field was also set at a constant level for all *S*s. The brightness of the exposure field was variable, always starting with an illumination well below that of the pre-exposure field and increasing by 1-v. steps up to the perceptual threshold. The tachistoscope was a double projector setup with synchronized camera shutters. Variacs wired into the projector circuit permitted manipulation of the voltage. The *S* sat in front of the projector at a distance of 8 ft. from the screen. The material was typed in Roman capitals with a sign typewriter and prepared as slides. Height of the letters projected on the screen was 17.5 cm. A slide in the projector for the pre-exposure field contained two black lines so placed as to frame the trigrams to be projected. Illumination of the projection room was low and always constant. Illumination of the pre-exposure field (set at 90 v.) measured 14 ft-L. When the background illumination of the exposure field was at 60 v. (a setting which permitted most of the material to be read) it measured 1.6 ft-L and a projected letter stroke measured .64.

The procedure for obtaining the threshold was an ascending method of limits, with *S* responding whenever he could by writing what he saw on a pad of paper on his chair arm. Each time he wrote something, *E* observed it and either increased the exposure illumination by 1 v. for another exposure, or told *S* he had read correctly and turned the page for recording responses to the next stimulus. After a successful trial, *E* read the voltmeter, reset it for trials with a new trigram, and gave *S* a ready signal. The instructions were as follows:

> This is an experiment on reading. We would like you to look at the screen in front of you, and when we say "ready" fixate the center of the two black lines. After the ready signal, a group of letters will be flashed on the screen between the lines for a brief interval. Please try to read it and write what you see on your response sheet. After writing, please look away from the screen until the next "ready" signal. At first, you will probably be unable to read anything. In that case, write nothing. We will continue with successive exposures until you have read a letter group correctly. As soon as you are fairly sure you have seen something, write it. But don't merely guess—write what you have seen only if you have a fair degree of confidence.

Three practice slides were given *S* (two 3-digit numbers and a word, BOB) during which the voltage suitable as a starting level for a given *S* was ascertained. The level was set so that *S* would be unable to report anything on the first exposure, but an attempt was made not to set it so low that any threshold measure would require more than about 15 steps. The 16 trigrams and words were then presented to *S*. Forty-two *S*s were run, yielding 14 threshold measures for each of the 36 trigrams. All *S*s were drawn from an introductory psychology course in the College of Arts and Sciences. None had rated the trigrams.

Method for testing recognition memory.—The same three sets of trigrams were presented in mixed lists of 12, again distributed so that only one member of a triad appeared in a list, but all appeared with equal frequency. As in the previous experiment, there were four trigrams of each type in a list. No real words were included. The three lists thus constituted were arranged in three orders with items rotated in each order so as to yield identical serial positions across the three members of a triad. The purpose of the three orders was to increase generality of any trend found, since counterbalancing for position was accomplished by maintaining identical positions for members of a triad.

The items were typed with the sign typewriter and photographed on film strip. A list was projected on a screen with a 2-sec. presentation for each item. Instructions were: "We are going to show you 12 three-letter combinations on this screen. Please notice them carefully." After the list had been exposed, *S* turned over a sheet of paper lying before him and read his further instructions, which were to read through all the letter combinations printed on the page, and check the 12 he had just seen. This page contained all the 36 trigrams, in random order. Three orders were again prepared so that position in the recognition test could be counterbalanced by assigning members of a triad to identical places. Each of the three test orders was used three times, a given test order always following a given exposure order so that no bias due to interaction of position of items in exposure list and test list was possible. The list-cum-test combinations were assigned at random to the nine subgroups, which were sections of an introductory psychology course. The *N*s for the counterpart-order subgroups were approximately equal. None of the 182 *S*s had taken part in the previous experiments.

Method for testing recall.—The film strips made for testing recognition were used for testing recall. However, the recall groups had three repetitions of a list. A different order appeared for each trial. As before, orders were held constant for members of a triad. At the end of each exposure of the list, *S* wrote all the items he could recall (in any order) on a pad of paper, turning back the page when he had finished. Again, the lists were projected with a 2-sec. presentation per item and 110 new *S*s were run in sections of the introductory course. Each of the three lists was given to two sections, randomly assigned.

RESULTS

Comparison of perceptual thresholds.—The threshold measures which had been recorded in volts were the settings required for an *S* to read a trigram correctly. As is the usual case with tachistoscopic recognition data, individual differences were very great. In order to determine whether there were further differential effects due to trigram type, the data were transformed to standard scores by the formula $50 \pm \left[10\,\frac{(x-\bar{x})}{\sigma}\right]$. The σ was determined for each *S* individually. Since each member of a triad was judged by different *S*s, the conversion to standard scores allowed statistical treatment of the differences between scores in individual triads, eliminating variability due to differences in legibility of letters across triads.

Although this procedure may have allowed slight differences in intra-*S* list composition to be obscured, such an effect would tend to work for the null hypothesis instead of against it; and since each *S* received four trigrams of each type (as well as four words) in his list, no one type of trigram would be favored by the standard score conversion.

The mean thresholds for each trigram, expressed as the converted scores, are given in Table 2. Comparison of the means at the bottom of the columns reveals that the overall mean threshold was lowest for the pronounceable trigrams, next lowest for the meaningful trigrams, and highest for the control items. The mean threshold for each trigram was treated as a score, and paired t tests between item types were calculated, using $df = 11$ for each test. Treating the data in this manner was more conservative than other possible treatments, but the t tests (two-tailed) were significant as follows: pronounceable vs. control, $p < .001$; pronounceable vs. meaningful, $p < .01$; meaningful vs. control, $p = .05$. In fact, the two larger differences would have been significant with a sign test. Eleven of the pronounceable trigrams had lower thresholds than their meaningful counterparts, and all 12 pronounceable ones were lower than

TABLE 2

Mean Thresholds, Recognition, and Retention (Recall) Scores for Triads of Meaningful, Pronounceable, and Control Trigrams

Meaningful				Pronounceable				Control			
Trigrams	Threshold	% Recog.	% Recall	Trigrams	Threshold	% Recog.	% Recall	Trigrams	Threshold	% Recog.	% Recall
RKO	45.2	94	68	KOR	41.6	64	57	OKR	43.7	61	46
TVA	47.2	92	76	TAV	46.5	74	18	AVT	46.6	51	22
IBM	45.8	95	51	MIB	42.8	74	25	MBI	48.4	56	22
FBI	57.6	97	73	BIF	53.0	92	49	IFB	55.4	61	43
ABC	48.3	98	83	BAC	44.9	50	51	ACB	52.0	80	46
TWA	52.6	95	83	TAW	51.4	42	44	ATW	65.3	39	11
ILR	52.0	97	83	RIL	47.0	71	39	RLI	56.2	62	15
WPA	57.9	84	43	WAP	50.5	80	51	APW	55.1	53	20
TKO	45.7	95	64	TOK	43.7	78	61	OTK	53.6	63	10
NYU	47.8	100	61	YUN	47.1	92	66	YNU	59.6	79	49
IFC	61.8	92	86	CIF	62.4	71	44	ICF	64.9	48	37
AFL	58.2	85	83	FAL	49.6	87	50	LFA	57.8	46	24
M	51.7	94	71.2		48.4	73	46.3		54.9	58	28.8
SD	5.6				5.4				6.4		

their control counterparts. Seven of the meaningful items had lower thresholds than their control counterparts.

It seems clear that pronunciability was more powerful in creating a unit of a three-letter group than was meaningfulness as defined in this experiment, if we accept lowness of perceptual threshold as an indicator of relative unitariness (processing as a whole rather than as discrete elements).

Four actual words were included among each *S*'s stimulus items. Thresholds for the words, since they were both meaningful and pronounceable, might be expected to be lowest of all the items. Such was the case (mean threshold = 45.0).

Analysis of prethreshold responses.—The prethreshold responses of all 42 *S*s were examined in an attempt to classify errors. Only three-letter responses were included. Two judges placed them in four categories: real words or names, initials or abbreviations, pronounceable trigrams, and trigrams which were not pronounceable and had no known significance. Words or names accounted for 32.5% of the errors, 4.7% of the errors were initials or abbreviations, 21.5% were pronounceable trigrams, and the rest were low in pronunciability and meaning. Although the probabilities for the various types of errors to occur are not known, there does not appear to be a set to look for initials during the perceptual task. This inference was corroborated by *S*s when questioned after the experiment.

The contribution of letter-shape confusions to erroneous responses is a matter of considerable importance for reading as well as for control of word-threshold experiments, but it is hard to evaluate quantitatively. One reason for the difficulty is the confounding of letter similarity and guessing from context. For example, the most frequent prethreshold response to the trigram BAC was BAG. Is this because BAG is a real word, and *S*, having perceived the first two letters, is simply operating on the generated frequency principle when he adds the G? Certainly this hypothesis is not entirely satisfactory, for other words such as BAR, BAT, and BAD are highly probable for this population, but were never emitted. On the other hand, C and G are two of the most confusable letters of the alphabet.

In an effort to assess the influence of confusions due to letter shapes, all the single-letter errors were collected in a confusion matrix. Only those errors which could be unequivocally located in the matrix were counted (no one- or two-letter responses were counted unless *S* had inserted dashes to make position in the presented letter group clear). The resulting matrix is not, of course, a proper one for studying letter confusability as such, but fortunately a matrix obtained under the most careful conditions with children, using the same type font, was available for comparison. Of the five most frequent letter confusions found in the prethreshold responses, four were among the frequent confusions made by the children (C with G, C with O, W with N, M with N). Examples of items yielding a number of these confusions are: ABO for ABC, AOB for ACB, ATN for ATW, NIB for MIB, YMU for YNU. When the structure of the trigram allowed, it seems clear that letter shape influenced *S*'s erroneous responses as well as his accurate ones.

In analyzing errors, a count was also made of errors in relation to position of the letter in the trigram. Errors were most frequent for the middle letter, slightly less for the third, and least for the first. Greater accuracy for the first letter is to be expected from *S*'s reading habits and from previous research (see Heron, 1957; Pillsbury, 1897). But accuracy on the first letter was by no means perfect. In considering the trigrams which had a higher frequency of errors on the first letter, and comparing them with their counterparts, the influence of letter shape is detectible. For instance, for the trigram CIF, errors were most on the first letter; for IFC, they were most on the third; and for ICF they were most on the second, with progressively greater frequency in the three cases. There appears, therefore, to be an interaction of position and letter shape.

Retention measured by recognition.—The percentage of *S*s that recognized each trigram correctly may be seen in Table 2. The strong trend of the threshold results, for pronounceable items to have an advantage over meaningful ones, .is reversed in the recognition scores. Meaningful trigrams are recognized correctly more often than both pronounceable and control ones (94% vs. 73% and 58%); pronounceable more often than control ones (73% vs. 58%). All the differences are significant by a sign test ($p < .01$, two-tailed).

The errors of recognition (marking an item as recognized when it was not a member of the list exposed to *S*) were analyzed to see if any of the three types of trigrams predominated. Most meaningful items were misrecognized (193 errors), control items next (173 errors), and pronounceable ones least (157 errors). Although these differences are not large, the trends were the same in the various subgroups of *S*s. Thus, meaning seems to have an advantage for recognition, measured by number of items correctly marked as recognized, but it is also associated with more frequent errors. Pronounceable items are recognized correctly more frequently than control ones, and are less often misrecognized. The higher frequency of checking the meaningful items, whether correct or incorrect, could therefore be due to a category set (looking for initials) which de-

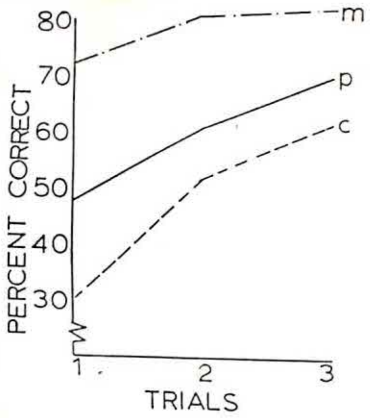


FIG. 1. Recall curves for three types of trigrams (Meaningful, Pronounceable, and Control).

veloped as *S* saw the list. But the advantage of pronounceable over control items (more chosen correctly and fewer chosen incorrectly) seems more likely due to their being perceived as units at the time of exposure. Structural cohesiveness due to pronunciability as a monosyllable gives *S* only one item to remember instead of three discrete letters.

Retention measured by recall.—The percentage of items recalled correctly on the first recall trial is presented in Table 2. It was highest for meaningful items, next highest for pronounceable items, and lowest for control items. The differences are significant by a sign test (two-tailed) as follows: meaningful vs. pronounceable, $p<.04$; meaningful vs. control, $p < .001$; pronounceable vs. control, $p < .01$. After the first recall trial, the differences between the three types of item decreased, as Fig. 1 shows. This trend is attributable to the fact that the number of meaningful items recalled correctly did not increase after Trial 2.

All the errors (incorrect items written in the recalls) were tabulated and examined for evidence of "coding" by category influence. That category membership (belonging to the class of "initials") is important in recall was suggested by the appearance of items such as YNU in a rearranged form, NYU, in the recalls. Evidence of a category set is also clear in importations such as FDR.

The errors are classified in Table 3. Two judges agreed on the classifications. A test of the differences between observed and expected number of rearrangements to the four categories yields a χ^2 of 65.7 with $df = 3$. This is significant at better than the .001 level.

It is interesting that there were far more rearrangements to initials than to any other category, including real words. There were 44% as many opportunities for word errors as for initial errors, but only 3% as many errors. The *S*s observed, as the list was exposed, that there were no real words but that there was a group of initials, and in recalling tried to make

TABLE 3

ANALYSIS OF RECALL ERRORS

Type of Error	Example	Observed	Expected
Rearranged to real word	CAB	1	4.6
Imported real words	SUP, GNU	12	
Rearranged to initial category	ICF to IFC	35	10.4
Imported initials	CIO	9	
Rearranged to pronounceable	RLI to RIL	19	27.6
Imported pronounceable	TOC	63	
Rearranged to control type	RLI to LRI	14	26.4
Others		99	

use of this knowledge. Of the 35 cases of rearrangement to the initial category, the item rearranged was more often a control item than a pronounceable item (28 to 7), suggesting that the pronounceable items were more resistant to rearrangement.

There were few rearrangements to pronounceable errors. With 265% as many opportunities for pronounceable errors as for initial errors, there were 54% as many errors. The pronounceable errors were not many more than were rearranged as the control type. With 105% as many opportunities, there were 136% as many pronounceable errors as control errors. The category of pronounceable monosyllables is no doubt too large to aid much in recall. Of the errors which were not real words or rearrangements of items exposed, a large number was pronounceable. They were often cases of one letter changed from the item exposed, such as CIV for CIF, TOC for TOK, and BAK for BAC. It was apparent from some of these errors that *S* had remembered the sound of a monosyllable and was misspelling it.

DISCUSSION

Two characteristics of verbal units—pronunciability and meaning—were compared in these experiments, and both were found to give an advantage to a sequence of letters for reading and remembering. But the striking fact which emerged is the discrepancy between the perceptual results and those for retention as regards the relative effectiveness of the two. Why should pronunciability be more effective here in structuring units for reading but meaningfulness be better for retention? Miller (1962) has pointed out that "language must contain some natural coding units of finite length. But what these units are for an articulate human being still remains something of a mystery [p. 81]." It is possible that the optimal coding units—or grouping principles—for one task are not optimal for another. An analysis of *S*'s task in the two situations may help.

In the perception of *written* language, the perceiver must code the stimulus material into units of *spoken* language—that is first and foremost the reading task. But we do not speak in letters; we speak in syllables, words, and larger functional units. A letter name is itself a morpheme, not an element of a speech unit. Since the reader must code the written letters into functional speech units, pronounceable letter combinations will obviously facilitate his task. But the kind of meaning present in IBM creates three units for reading rather than one.

In the recognition and recall tasks, on the other hand, the stimulus items have already been read (coded into speech units). The task now is to hang onto them—store them for future "retrieval." Much recent research has demonstrated that coding for retrieval is most efficient when the items are subsumable under a known and "exhaustible" category (see Cohen, 1963). The category of pronounceable trigrams (if it could be called a category at all) is so large as to be of small help in retrieval. In other words, pronunciability is of the utmost help in structuring a unit—grouping the letters into a single item—but it is of little help in grouping these items under a single address. When larger linguistic structures are possible, such as phrases or sentences, they obviously operate to facilitate retention. But linguistic structure can help no farther than the single syllable unit in the present experimental situation.

On the other hand, the class of "well-known initials" does provide a common category for grouping items in storage. It provides a coding principle for retention, and that *S*s used this principle is clear. In recognition, more of the meaningful items were checked than any others, whether right or wrong. And in recall, *S*s demonstrated use of the initial category by errors of rearrangement. Underwood and Keppel (1963) found that if an *S* "encodes" a nonsense syllable into a word and must "decode" it again in recall to the original syllable, performance is inhibited, rather than facilitated. In the present experiment, failure in decoding can be inferred to be responsible for the errors of rearrangement in recognition and recall. The parallel facilitation in both recognition and recall can then be inferred to be the result of grouping the "initial" items actually presented into one category, from which they were retrieved at the time of recognition or recall.

Consider now the question of response "integration." Meaningful items possessing high "generated frequency" seem to exemplify, maximally, integration by associative frequency. But one cannot conclude that the better retention of these items over pronounceable ones proves the greater effectiveness of associative integration. These items did not have much advantage over control ones as units in perception, and their advantage over pronounceable ones in retention can be accounted for by *S*'s tendency to group them into a useful category. It was thus not only their high letter-sequence associations but the possibility of grouping them under one heading that made the difference. It is interesting to note in this connection that nonsense syllables (paired associates) are harder to learn when they are spelled out than when they are pronounced as units (Rock & Steinfield, 1963). It is possible that initial responses are integrated by associations between the letters but the present results can be explained plausibly without such a principle. That pronunciability confers unity is supported.

REFERENCES

Bryan, W. L., & Harter, N. Studies on the telegraphic language. *Psychol. Rev.*, 1899, **6**, 345–375.

Cattell, J. McK. Ueber die Zeit der Erkennung und Benennung von Schriftzeichen Bildern und Farben. *Phil. Stud.*, 1885, **2**, 635–650.

Cohen, B. H. An investigation of recoding in free recall. *J. exp. Psychol.*, 1963, **65**, 368–376.

Erdmann, B., & Dodge, R. *Psychologische Untersuchungen über das Lesen.* Halle: Niemeyer, 1898.

Gibson, E. J., Osser, H., & Pick, A. D. A study of the development of grapheme-phoneme correspondence. *J. verbal Learn. verbal Behav.*, 1963, **2**, 142–146.

Gibson, E. J., Pick, A. D., Osser, H., & Hammond, M. The role of grapheme-phoneme correspondence in the perception of words. *Amer. J. Psychol.*, 1962, **75**, 554–570.

Heron, W. Perception as a function of retinal locus and attention. *Amer. J. Psychol.*, 1957, **70**, 38–48.

Mandler, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.

Mandler, G. From association to structure. *Psychol. Rev.*, 1962, **69**, 415–427.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

Miller, G. A. Decision units in the perception of speech. *IRE Trans. inform. Theory*, 1962, **8**, 81–83.

Noble, C. E. Meaningfulness and familiarity. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963. Pp. 76–119.

Pillsbury, W. B. The reading of words: A study in apperception. *Amer. J. Psychol.*, 1897, **8**, 315–393.

Rock, I., & Steinfield, G. Methodological questions in the study of one-trial learning. *Science*, 1963, **140**, 1–2.

Underwood, B. J., & Keppel, G. Coding processes in verbal learning. *J. verbal Learn. verbal Behav.*, 1963, **1**, 250–257.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

(Received March 16, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 183–190

ACQUISITION AND RETENTION OF CONSISTENT ASSOCIATIVE RESPONSES[1]

LEO POSTMAN

University of California, Berkeley

The effects of pre-experimental language habits on the acquisition and retention of consistent associative responses to word stimuli were investigated. A variant on the standard procedure of paired-associate learning was used in which S is required to provide his own responses to each of a series of stimuli and to use these responses consistently thereafter. Words of high, medium, and low frequency of usage were used as stimuli. Practice was continued to a criterion of 3 trials of consistent responding, and retention of the terminal associations was tested 30 sec. or 48 hr. after learning. Speed of acquisition 1st increased and then decreased as a function of stimulus frequency. Retention over a 48-hr. period was high but varied inversely with stimulus frequency. The results provide evidence for correlated increases in positive and negative transfer effects of language habits as a function of the meaningfulness of the stimulus terms.

When the materials in a rote-learning experiment consist of words in *S*'s language, pre-experimental associations elicited by the items in the list may serve as sources of both facilitation and interference. To the extent that the prescribed associations conform to the learner's language habits, unit-sequence facilitation will result. As the new associations diverge from prior language habits, unit-sequence interference will develop. Both unit-sequence facilitation and interference may be expected to increase with the meaningfulness of the items in the list. The larger the number of different associations which an item acquires through linguistic usage the more readily it can be linked with other items, either directly or through short mediational chains. At the same time, however, the amount of interference during acquisition and retention may be expected to increase with the number of pre-experimental associations which can compete with the prescribed connections. These dual and opposed effects of pre-experimental habits have been described as the interference paradox of associative probability (Underwood & Schulz, 1960, p. 46).

In the acquisition and recall of paired associates, correlated increases of unit-sequence facilitation and interference should occur as a function of the meaningfulness of both stimulus and response terms. The relative amount of interference may, however, be expected to grow more rapidly with increases in stimulus than response meaningfulness. Since paired items are likely to be practiced in the order of their presentation, competing associations elicited by the stimuli should be more effective in interfering with learning and recall than those evoked by the response terms. In accordance with these expectations, it was found in a previous study (Postman, 1962) that speed of learning first increased and then decreased as a function of stimulus meaningfulness but varied directly with response meaningfulness. Recall after 1 wk. was not influenced by response meaningfulness whereas the effects of stimulus meaningfulness paralleled those in acquisition. It was concluded that long-term recall was significantly influenced by the recovery of competing associations elicited by the stimulus terms.

---

[1] This research was supported by a grant from the National Science Foundation and Contract Nonr 222 (90) between the University of California and the Office of Naval Research.

When a specific response is prescribed for each of the stimuli in a list, as is the case in standard paired-associate learning, the number of pre-experimental language habits which are activated in the course of learning is likely to be limited. The range of responses emitted by *S* during the acquisition of a rote series is severely restricted by the operation of a "selector mechanism" (Underwood & Schulz, 1960, pp. 143 f.), i.e., *S* consistently confines his responses to items from within the list. As a result, facilitation and interference during acquisition and recall will be determined largely by pre-existing associations among the items in the list. The potential influence of pre-experimental language habits becomes less restricted in a situation in which *S* is free to generate his own responses to a set of stimuli, with the requirement that the responses be used consistently from trial to trial. Such a procedure was first used by Mattocks (described by Underwood & Schulz, 1960, pp. 273–278). In that study *S*s were instructed to give three-letter words as responses to each of a series of trigrams of low meaningfulness; practice was continued to two successive perfect trials. In the absence of pre-experimental associations to the stimuli the task proved to be difficult. Although *S*s' responses tended to be of high frequency of usage, the associations to the stimuli of low meaningfulness were quite unstable and were changed repeatedly in the course of practice. By contrast, it was found in a subsequent study by Postman and Stark (1962) that consistent sets of responses were generated very rapidly when the stimuli were English words.

The present study investigates the acquisition and retention of a set of consistent associative responses as a function of the meaningfulness of the stimulus words. The procedure developed by Mattocks, which will be designated as the Method of Generated Responses (MGR), was used. Our analysis of the role of associative probability leads to the expectation of dual and opposed effects of increases in stimulus meaningfulness. Since strength of pre-experimental associations and amount of response competition are assumed to be correlated, an inversion in the relationship between meaningfulness and performance is expected in both acquisition and retention.

METHOD

Materials.—The stimulus series consisted of 12 two-syllable nouns. There were two lists each of words of high, medium, and low frequency of usage (Lists HF, MF, and LF). These lists were drawn from three frequency ranges in the "L" count of Thorndike and Lorge (1944): 1,000–3,300, 10–33, and 1–3 occurrences in 4.5 million. The mean *m* values and rated familiarity of the items, which were obtained in prior standardization procedures, are correlated with word frequency. The words were chosen so as to minimize pre-experimental associations among the items in each list. Norms of free association, based on the responses of 1,000 students at the University of California, were used for this purpose. There was no duplication of first letters in any of the lists. The two lists at each frequency level were used equally often.

Procedure.—In MGR the experimental arrangements are the same as in standard paired-associate learning except that the response side of each "pair" is left blank. A 2:2 rate of presentation was used, with an intertrial interval of 8 sec. The rise of the shutter after 2 sec. signaled the end of the response interval, just as it does in the standard procedure. Thus, *S* was exposed to the stimulus term for 4 sec., but was required to respond during the first 2 sec. of the cycle. The *Ss* were informed that they had to decide on a different response to each of the stimulus terms and that they were to continue giving the responses which they had chosen. While it was permissible to change responses, practice would be continued until a predetermined level of consistent responding had been achieved. No correction was given during practice.

Five different orders of the stimuli were used to minimize serial learning. Each of these orders was used as a starting order equally often. Practice was continued to a criterion of three successive trials of consistent responding.

Retention was tested by recall and relearning either 30 sec. or 48 hr. after the end of original learning. The *Ss* were instructed to try to give the responses which they had used on the criterial trials in original learning. The *Ss* were reminded of the procedure used in acquisition and were urged to recall the last response they had given to each of the stimulus words. Again no correction was given. Relearning was continued to a criterion of three consistent trials or for five trials, whichever took the longer.

Subjects.—With three kinds of lists and two retention intervals, there were six groups of 20 *Ss* each. The *Ss* were undergraduate students who had English as their native language. They had no prior experience with MGR but were not necessarily naive to rote-learning experiments. The *Ss* were assigned to conditions in blocks of 6, with 1 *S* per block for each combination of lists and retention intervals. The running order within each block and the assignment to specific lists and running orders were determined by means of a table of random numbers, with restriction to equal *N*s. No *Ss* were lost because of failure to learn.

RESULTS

Acquisition.—The 30-sec. and 7-day groups did not differ significantly in original learning. For the combined groups, the mean numbers of trials to the criterion of three consistent repetitions were 8.38 ($SD = 4.44$) for List HF, 6.10 ($SD = 2.21$) for List MF, and 7.00 ($SD = 2.95$) for List LF. Speed of mastery first increased and then decreased as a function of the word frequency of the stimulus terms. The effect of word frequency is significant beyond the .01 level, $F\ (2, 117) = 5.00$, after log transformation to remove heterogeneity of variance.

The mean trial of the first associative response to the stimulus terms was 1.26 ($SD = .25$) for List HF, 1.36 ($SD = .39$) for List MF, and 1.39 ($SD = .45$) for List LF. The corresponding values for the first occurrence of the association used on the criterial trials were 1.50 ($SD = .40$), 1.53 ($SD = .63$), and 1.58 ($SD = .84$). The three types of lists do not differ significantly by either measure, although it should be noted that the variability increases as the word frequency decreases. The absence of substantial differences in the time of initial occurrence of associations indicates that speed of mastery was determined primarily by the amount of oscillation from trial to trial, i.e., the stability of the associative responses. An analysis of the errors during acquisition supports this conclusion.

Errors during acquisition.—The errors made on the acquisition trials were classified into three categories: (*a*) failures to respond, (*b*) intralist errors, and (*c*) drop-out errors. For purposes of this classification the responses on the criterial trials were considered as analogous to correct responses in standard paired-associate learning. On this basis an intralist error for a given stimulus term was defined as an association which was not used on the criterial trials and which occurred as a response to another stimulus at least once. Responses which were given to only one stimulus term and which were not retained on the criterial trials were classified as drop-out errors. Table 1

TABLE 1

MEAN NUMBER OF OCCURRENCES PER TRIAL OF DIFFERENT TYPES OF ERRORS

List	Failures to Respond	Intralist	Drop-Out	% Overt Errors
HF	1.17	.36	.18	35.9
MF	1.54	.17	.18	21.3
LF	1.55	.24	.12	22.9

shows the mean numbers of occurrences per trial for each class of errors.

Failures to respond tended to occur less often for List HF than for Lists MF and LF, but the differences among the groups are not significant. The rate of intralist errors first increases and then decreases as a function of word frequency, just as do the numbers of trials to criterion. The variation in the numbers of intralist errors per trial is significant by the median test which was applied in view of the skewness of the distributions, $\chi^2\ (2) = 7.82$, $p < .02$. Drop-out errors were more likely during the acquisition of Lists HF and MF than of List LF but the differences are not significant. Thus, there is a reliable effect of word frequency only on the rate of occurrence of intralist errors. The rank orders of the conditions for the different classes of errors indicate that a larger proportion of all errors were overt responses for List HF than for the other two lists. The mean percentages that intralist and drop-out errors were of all errors are shown in Table 1. There is a clear separation between List HF on the one hand and Lists MF and LF on the other. The variation in percentages is significant by the median test, $\chi^2\ (2) = 10.07$, $p < .01$.

Distributions of associations.—Since the course of learning in MGR is assumed to exhibit the influence of pre-experimental habits, it becomes important to determine to what extent the responses in acquisition reflect *S*s' habitual associative hierarchies. The distributions of responses in MGR were evaluated against the norms of free association for the stimulus terms. Table 2 presents the mean percentages of occurrence of different first responses in MGR as a function of the rank of these responses in free association. The normative percentages are included for purposes of comparison. The MGR distributions are based on the responses of 40 *S*s whereas the norms of free association were obtained from 1,000 *S*s.

The distributions of responses in MGR agree quite well with the

TABLE 2

RANK-FREQUENCY FUNCTIONS OF FIRST RESPONSES IN MGR AND OF FREE ASSOCIATIONS TO STIMULUS TERMS

Rank	List HF		List MF		List LF	
	MGR	FA	MGR	FA	MGR	FA
1	28.3	26.5	31.4	25.3	32.7	23.8
2	15.0	10.7	12.7	11.9	10.0	12.8
3	6.7	6.3	5.2	7.3	7.5	7.1
4	7.5	5.1	4.8	5.2	4.6	4.8
5	2.9	4.0	1.9	3.5	1.9	3.4
>5	29.1	47.3	32.1	46.8	28.3	48.1
New[a]	10.4		11.9		15.0	

[a] Not in tables of association norms.

normative rank-frequency functions. For each of the three lists, however, the percentage of primaries is higher in MGR than in the norms. This discrepancy is more pronounced for Lists MF and LF than for List HF. For the secondary response only List HF shows a clear discrepancy in the same direction. Thus, the bias toward associates of high rank appears to be present in List HF but influences the percentage of secondaries more than that of primaries.

It is likely that *S*s in MGR favored associates of high rank because the apparent strong connection between such responses and the stimulus terms might be expected to aid retention. If such a strategy was, indeed, pursued, it probably had only limited effects on the speed of acquisition. The correlation between the number of primaries given by *S* as first associations and trials to criterion was −.08 for List HF, −.04 for List MF, and −.36 ($p < .05$) for List LF. Thus, the use of primaries facilitated learning to a moderate degree only for List LF. This finding is in agreement with the results of an earlier study (Postman, 1962) which showed that the associative rank of the response terms influences speed of paired-associate learning when the word frequency of the stimulus terms is low but not when it is high.

The changes between the first associations and those used on the criterial trials were classified according to the direction of change in normative rank as (*a*) upward shifts to a response of higher probability, (*b*) downward shifts to a response of lower probability, and (*c*) invariant in probability. The numbers of upward shifts were 21, 12, and 12 for Lists HF, MF, and LF, respectively. The corresponding numbers of downward shifts were 19, 10, and 15. The numbers of invariant cases were 0, 3, and 7. The occurrences of upward and downward shifts were approximately balanced. These terminal shifts represent, of course, only a part of the associative changes during learning. More than two responses were given to some stimuli; in other cases the initial response was displaced temporarily and reinstated on the criterial trials. However, as the terminal shifts indicate, there was no systematic trend toward a change in the ranks of the associates.

Word frequency of responses.—While the rank-frequency functions for the three lists were comparable, there were clear differences in the word frequency of the responses given to the three sets of stimulus terms. The mean log frequency of occurrence of the terminal response words in the "L" count of Thorndike and Lorge (1944) was determined for each *S*. The means of these distributions were 2.84 ($SD = .25$) for List HF, 2.42 ($SD = .36$) for List MF, and 2.25 ($SD = .34$) for List LF. The differences among groups are significant beyond the .01 level, $F(2,117) = 13.44$. The relationship between the word frequency of the stimuli and of the responses is direct whereas speed of acquisition first increases and then decreases with stimulus frequency. It is apparent, therefore, that the variations in rate of learning are not a simple function of the word frequency of the responses and are significantly influenced by the associative probabilities of the stimulus terms. It is not possible to say whether the word frequency of response and stimulus terms interacted to determine the course of acquisition. The absence of significant correlations within groups between word frequency of responses and number of trials to criterion indicates, however, that there was

TABLE 3

MEAN NUMBERS OF RESPONSES RECALLED AND MEAN NUMBERS OF TRIALS IN RELEARNING TO CRITERION

List	30-Sec. Test				48-Hr. Test			
	Recall		Trials in RL		Recall		Trials in RL	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HF	11.85	.36	3.55	1.36	11.20	.98	4.70	1.93
MF	11.95	.22	3.45	.97	11.55	.59	3.45	.59
LF	11.95	.22	3.20	.60	11.85	.36	3.35	1.41

probably no such interaction. These correlations were .13, .21, and −.07 for Lists HF, MF, and LF, respectively. The word frequency of the responses was high for all lists, and variations within the upper range of frequencies are not likely to have important effects on acquisition.

Recall.—The mean numbers of criterial responses recalled after 30 sec. and 48 hr. are shown in Table 3. The most striking feature of these results is the extremely high level of retention. Recall is virtually perfect after 30 sec. There are small losses after 48 hr. which vary inversely with stimulus frequency. The largest loss, which is for List HF, is less than one item. Although the differences in forgetting are quite small, they are statistically significant. Analysis of variance of the recall scores (following a Freeman-Tukey square-root transformation) yields $F\ (2, 117) = 3.08$, $p < .05$. The majority of the errors at recall were failures to respond. After 48 hr., these account for $\frac{13}{16}$, $\frac{7}{9}$, and $\frac{2}{3}$ of the items forgotten from Lists HF, MF, and LF, respectively.

The probability of a criterial response being forgotten should vary with the stability of the association during acquisition. For purposes of assessing this relationship, the items learned by each *S* were divided into two categories: (*a*) stable associations for which the response remained unchanged from the initial to the criterial trial, and (*b*) unstable associations for which associative shifts and omissions occurred during practice. The items were pooled for all *S*s tested on a given type of list after 48 hr. For List HF, 4.3% of the stable items and 11.4% of the unstable items were forgotten. The corresponding percentages for List MF were 2.0 and 5.4; for List LF they were 0.6 and 3.1. In each case the percentage of losses is lower for stable than for unstable items. This relationship parallels that between number of reinforcements and retention in conventional studies of rote learning.

Relearning.—The mean numbers of trials in relearning to a criterion of three trials of consistent responding are shown in Table 2. Relearning was extremely rapid, and the differences among conditions are again quite small. An inverse relationship between word frequency and speed of relearning is, however, found on both retention tests. The overall effect of word frequency is significant, $F\ (2, 117) = 5.55$, $p < .01$, after log transformation. The interaction of Word Frequency with Time Interval falls short of significance, $F\ (2, 117) = 2.85$.

Discussion

The speed with which consistent associative responses are generated by *S* first increases and then decreases as a function of the word frequency of the stimulus terms. This relationship parallels that obtained in standard paired-associate learning (Postman, 1962) and is assumed to reflect the dual and opposed effects of associative probability. According to this interpretation, the absolute strength of pre-experimental associations increases with word frequency but so does the amount of competition among alternative responses. The U shaped relationship between word frequency and speed of acquisition is consistent with the assumption that unit-sequence interference increases at a faster rate than unit-sequence facilitation.

While the rate of acquisition in MGR is more rapid than in the conventional procedure, there are important similarities between the characteristics of performance in the two situations. Although the responses were not prescribed and no reinforcements were given in MGR, *S*s tended to confine their responses largely to their initial associations. As a result, shifts in the stimuli eliciting these initial responses were a major source of errors. Such shifts are equivalent to misplaced responses in standard paired-associate learning. Thus, the selector mechanism described by Underwood and Schulz (1960, pp. 143 f.)—*S*'s tendency to limit his overt responses to items within the list—appears to operate in MGR. The instructions to respond consistently probably led *S*s to reject responses which they recognized as new.

After the criterion of consistent responding had been reached, retention over a 48-hr. period was extremely high. This finding suggests that interference at recall was effectively reduced by differentiation, i.e., *S*'s success in discriminating between correct and incorrect responses. Degree of differentiation is assumed to vary directly with the absolute strengths of the competing responses as well as with the difference in their strengths (Underwood, 1945). At all levels of word frequency the associations generated in MGR are, by the very nature of the experimental situation, relatively high in the learner's hierarchy. Nevertheless, as the errors in acquisition indicate, *S*s' initial responses are subject to competition from alternative associations. Once a set of high-strength associates has achieved dominance, however, there should be maximal differentiation at recall. The fact that most of the errors on the test of recall were failures to respond indicates that *S*s were, indeed, successful in discriminating between correct and incorrect associations. With highly available responses the rejection of an error may well allow sufficient time for the substitution of the correct association.

The small amounts of forgetting which did occur were inversely related to the word frequency of the stimulus terms. These losses are attributable to the recovery of competing associations and failures of differentiation. With practice carried to a high criterion, initial differences in unit-sequence facilitation are minimized, and the retention losses reflect directly the increases in unit-sequence interference as a function of word frequency. These results are consistent with the earlier finding that high-probability associates learned under standard conditions are retained better when the word frequency of the stimuli is low than when it is high (Postman, 1962).

Two features of the present results should be emphasized in conclusion: (*a*) When *S*s are free to choose their own responses to stimulus words varying widely in frequency of usage, the pattern of differences in acquisition and recall is consistent with the assumption of correlated increases in unit-sequence facilitation and interference as a function of meaningfulness. Thus, the present results support the interpretation of earlier findings obtained in standard rote-learning situations. (*b*) The pattern of errors in learning and recall as well as the uniformly high level of retention add to the mounting evidence (cf. Postman,

1963) for the effectiveness of the selector mechanism in limiting the range of alternative responses and of the mechanism of differentiation in reducing interference in recall.

REFERENCES

Postman, L. The effects of language habits on the acquisition and retention of verbal associations. *J. exp. Psychol.*, 1962, **64**, 7–19.

Postman, L. Does interference theory predict too much forgetting? *J. verbal Learn. verbal Behav.*, 1963, **2**, 40–48.

Postman, L., & Stark, K. Retroactive inhibition as a function of set during the interpolated task. *J. verbal Learn. verbal Behav.*, 1962, **1**, 304–311.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**(3, Whole No. 273).

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received March 20, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 191–192

SUPPLEMENTARY REPORTS

TEMPORAL INTERACTION BETWEEN ELECTROCUTANEOUS STIMULI[1]

BURTON S. ROSNER

School of Medicine, Yale University

The effect of a rectangular electrocutaneous conditioning shock upon the absolute threshold for a subsequent test shock to the same locus was measured by the method of limits. Time between conditioning and test stimuli and intensity of the former were parameters. The conditioning shock raises the threshold for the test shock; this "inhibitory" effect increases with stronger conditioning shocks and with shorter times between the 2 stimuli. The effect is more striking than similar inhibitory phenomena which occur when conditioning and test stimuli impinge upon different cutaneous loci. These findings suggest one mechanism for explaining Békésy's observation that subjective vibratory pitch may drop with increased frequency or intensity of repetitive stimulation.

Schmid (1961) reported that a conditioning shock to one finger raises the absolute threshold for a test shock to another finger on the same hand. The "inhibitory" effect decreases as intensity of the conditioning shock decreases and as time or distance between the two stimuli increases. Uttal (1960) found a somewhat similar effect, upon varying distance between stimuli. The present report describes a limiting case of Schmid's experiment, where both stimuli are delivered to the same finger. This spatial arrangement of stimuli should produce a very strong inhibitory effect.

Method.—Schmid has described electrodes and stimulating equipment. For this experiment, conditioning and test stimuli were delivered to the two most distal phalanges of the left index finger; the cathode was proximal. A small box with a row of three push-button switches and a yellow and green light lay at *S*'s right hand. The green light was on at the start of each trial. When *S* was ready, he pushed the middle switch. He thereby turned off the green light, turned on the yellow one, and then received a stimulus 2.5 sec. later. The *S* signaled that he did or did not feel the test shock by depressing, respectively, the left-hand or right-hand switch. After recording *S*'s response, *E* set the intensity of the test shock for the next trial and then turned off the yellow light and turned on the green one.

The method of limits was used to obtain absolute thresholds for the test shock alone and for the test shock preceded by a fixed conditioning shock. Each experimental session consisted of eight ascending and eight descending series in counterbalanced order; the conditioning shock occurred on even-numbered series only. Intensity (4.0, 8.0, or 16.0 ma.) of the conditioning shock and time between conditioning and subsequent test shocks (30, 50, 75, 100, 125, 150, 175, 200, 250, or 500 msec.) were fixed for each session. Different combinations of intensity and intershock interval (Δt) were used for different sessions. Preliminary observations were made on six *S*s, of whom two were selected for extensive studies. These two *S*s received 12 additional practice sessions before the main experiment.

Results and discussion.—Preliminary experimentation confirmed the presence of a strong inhibitory effect in all *S*s. Results for the two *S*s used in the main experiment appear in Fig. 1 and 2. The ordinate in each figure is the difference between the threshold for the test shock preceded by the conditioning stimulus (I_{TC}) minus the threshold for the test shock alone (I_{TA}). The abscissa is time between conditioning and test shock (Δt) and the parameter is intensity of the conditioning shock (I_C).

The absolute threshold for Subject TA (Fig. 1) was 1.97 ± .22 ma. over the course of the experiment. This *S* showed no difference between ascending and descending determinations of I_{TA} but did show a statistically significant ($p < .01$ by Wilcoxon's matched pairs test) error of anticipation of 0.02 ma. in determinations of I_{TC}. Since this was about 1% of his absolute threshold, which always was equal to or below I_{TC}, data for ascending

---

[1] This work was supported by Grants M-1530 and M-5286 from the National Institute of Mental Health.

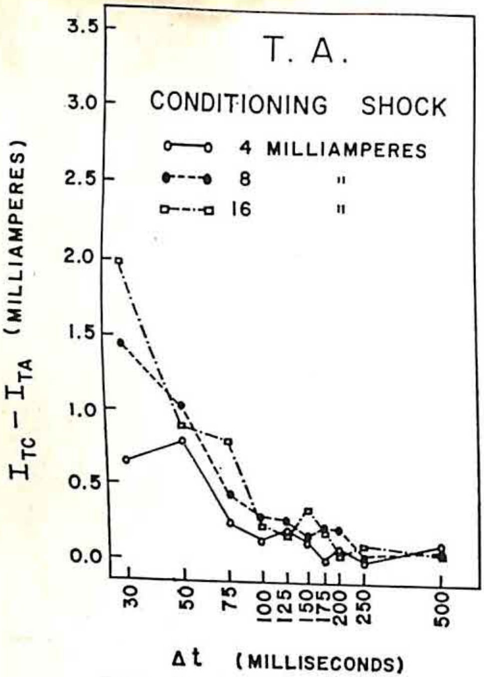


FIG. 1. Results for Subject TA. (Explanation in text.)

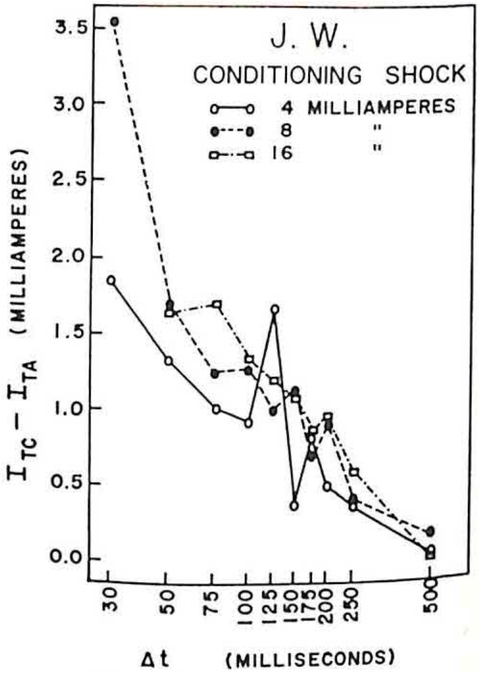


FIG. 2. Results for Subject JW.

and descending series were combined for further analysis. Application of the Mann-Whitney *U* test showed that I_{TC} exceeded I_{TA} at all intensities up to a t of 150 msec. ($p < .01$). At larger values of Δt, there were scattered instances of significantly elevated values of I_{TC}. The magnitude of the effect at these delays, however, was quite small. At intervals below 30 msec., *S* could not discriminate two shocks as discrete stimuli.

Subject JW (Fig. 2) had an absolute threshold of 1.55 ± .14 ma. and showed no systematic errors of habituation or anticipation. For this *S*, I_{TC} exceeded I_{TA} at a .01 level of confidence or better for all values of Δt up to 500 msec. The differences were not significant at this longest delay. This *S* could not discriminate between two 16-ma. shocks at a 30-msec. Δt or between two less intense shocks at intervals below 30 msec.

The rise in the absolute threshold for a test shock induced by a conditioning shock to the same locus is larger and more enduring than that found by Schmid for spatially diverse stimuli. The phenomenon drops out as time between stimuli increases; apparently, the slope of this decrease varies between *S*s. Increased intensity of the conditioning stimulus heightens the inhibitory effect at short Δt's.

These findings confirm and extend previous reports of temporal interactions in somesthesis. They also may provide some insight into Békésy's (1962) finding of a drop in vibratory pitch with increased intensity or even increased frequency of a repetitive stimulus. The inhibitory interaction between two successive stimuli reported here should produce subjective "drop-outs" of some members of a train of rapidly recurring stimuli. The rate of drop-outs should increase with intensity or frequency of stimulation. Since recovery of human peripheral nerve follows a much faster time course than psychophysical interaction between two stimuli (Rosner, 1961), this interaction probably is predominately central in origin. Reports of prolonged recovery cycles in human central neural evoked responses to somatic stimuli (Allison, 1962; Shagass & Schwartz, 1961) provide some confirmation for this contention.

REFERENCES

ALLISON, T. Recovery functions of somatosensory evoked responses in man. *EEG clin. Neurophysiol.*, 1962, **14**, 331–343.

BÉKÉSY, G. VON. Can we feel the nervous discharges of the end organs during vibratory stimulation of the skin? *J. Acoust. Soc. Amer.*, 1962, **34**, 850–856.

ROSNER, B. S. Neural factors limiting cutaneous spatiotemporal discriminations. In W. A. Rosenblith (Ed.), *Sensory communication.* Cambridge, Mass.: Technology Press, 1961. Pp. 725–737.

SCHMID, E. Temporal aspects of cutaneous interaction with two-point electrical stimulation. *J. exp. Psychol.*, 1961, **61**, 400–409.

SHAGASS, C., & SCHWARTZ, M. Reactivity cycle of somatosensory cortex in humans with and without psychiatric disorder. *Science*, 1961, **134**, 1757–1759.

UTTAL, W. R. Inhibitory interaction of responses to electrical stimuli in the fingers. *J. comp. physiol. Psychol.*, 1960, **63**, 47–51.

(Received March 2, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 193–195

HUMAN HEART RATE RESPONSES DURING EXPERIMENTALLY INDUCED ANXIETY: A FOLLOW UP WITH CONTROLLED RESPIRATION[1]

GEORGE E. DEANE

Harpur College

$\frac{1}{2}$ of the Ss were allowed to breathe normally and $\frac{1}{2}$ were required to breathe at a constant rate throughout. All Ss were then told to expect shock during No. 10 in a sequence of No. 1–12. $\frac{1}{2}$ of the Ss received one shock. All Ss showed an acceleration in heart rate during No. 1 through 8 and a deceleration during No. 10.

Previous studies by Deane (1961) and Jenks and Deane (1963) have provided evidence that there are two opposing heart rate responses during experimentally induced anxiety. Since Westcott and Huttenlocher (1961) have shown that extensive and reliable effects on heart rate may be attributable to respiratory activity, it was felt that an attempt should be made to determine whether or not the cardiac rate changes reported by Deane and by Jenks and Deane were a function of differential respiratory activity. The Westcott and Huttenlocher findings suggest that the acceleration noted in the previous studies might have resulted from more rapid and deep breathing during the shock-anticipation trials and that the deceleration might have been the result of a sharp inspiration followed by a continuous expiration during the time the noxious stimulus was anticipated. In the present study half of the *S*s were allowed to breathe normally as in the earlier studies but with respiration recorded, while the other half were required to breathe at a constant rate throughout. If the cardiac activity in the controlled-breathing group differs from that of the normal-breathing group, then the respiration records of the normal-breathing group should reveal the specific type of respiratory activity associated with the cardiac changes which occur.

Method.—The *S*s were 40 male students at Harpur College. A Grass Model 5 polygraph in conjunction with a chest bellows and Grass Model CP1 transducer were used to record respiration. A Grass Model 5P4 preamplifier was used for EKG recording from Standard Lead III. Chart speed was 25 mm/sec. Green and red jewel lights mounted on the face plate of the memory drum were controlled by Hunter timers to go on alternately for 1.8 sec. each to pace the inhale-exhale cycle. Stimulus numbers were presented at the rate of one every 3 sec.

After the chest bellows and cardiograph leads were attached, *S*s in the controlled-breathing groups were given 5 min. of practice inhaling while the green light was on and exhaling while the red light was on. The *S*s in the normal-breathing groups were allowed an equal amount of time to breathe normally. Base-level measurements were made as in the earlier study except that *S*s in the controlled-breathing groups were instructed to synchronize their breathing with the lights when they appeared, and to breathe normally at all other times. The respiration control lights were turned on by *E* on each trial sometime during the presentation of No. 1, 2, or 3. At the end of the tenth base-level presentation each *S* was told that he would continue to see the series of No. 1 through 12 but that now when the number 10 appeared he would receive a shock on the first two fingers of his right hand. Actually, in the controlled-breathing group 10 *S*s (Group C_1) received a shock on the first of the following six trials while the other 10 *S*s (Group C_0) never received shock. In the normal-breathing group the 10 *S*s in Group N_1 received one shock while the 10 *S*s in Group N_0 received none.

Results and discussion.—In Fig. 1 the results for Groups C_0 and C_1 combined are indicated by the solid line and the results for Groups N_0 and N_1 combined are indicated by the dashed line. The results were combined in this manner for Fig. 1 since the differences between the two controlled-breathing groups and between the normal-breathing groups were not significant. For both groups the line with the filled circles is a mean of the last three base-level presentations (B8–10), the line with the open circles is a mean of all six shock-anticipation trials (T1–6), and the line with the open squares is a mean of the last five shock-anticipation trials (T2–6). At Stimulus No. 1 → 8 is plotted the mean cardiac rate during these numbers, and at

---

[1] This research was supported in part by Grant MH-06590-01 from the National Institute of Health, Public Health Service. The author wishes to express thanks to Donald K. Stammer for his assistance in the collection and analysis of data.

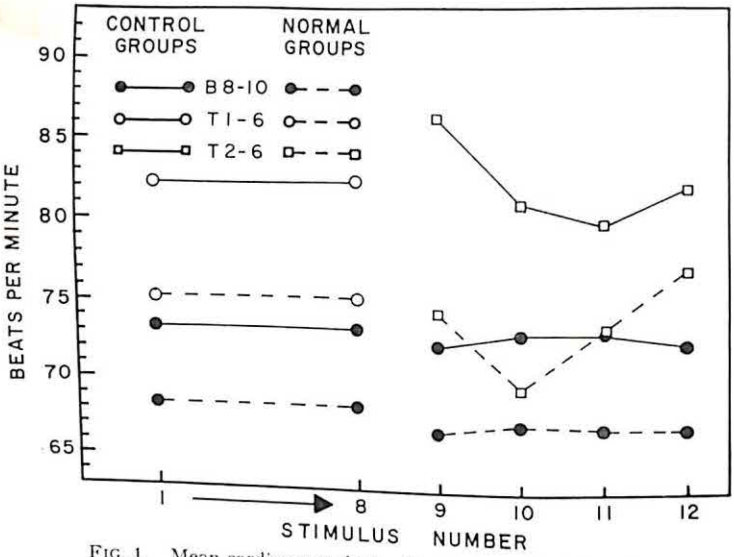


FIG. 1. Mean cardiac rate during the number stimuli for the controlled-breathing and normal-breathing groups.

each of the points labeled Stimulus No. 9 through 12 is plotted the mean cardiac rate of the two middle beats occurring during presentation of each of these numbers.

Figure 1 indicates that for both groups the effect of instructions regarding impending shock was to accelerate the rate over the base level during the first eight stimulus numbers. An analysis of variance was performed on the mean rate during No. 1 through 8 for Trials B8–10 and T1–6, and Table 1 shows the Trials effect to be significant at better than the .001 level. The nonsignificant Trials × Instructions interaction indicates that the magnitude of acceleration was as great for those *S*s required to control their rate of breathing as for those *S*s allowed to breathe normally. Figure 1 also shows that for both groups there was a deceleration in rate on Trials T2–6 during No. 10 where shock was anticipated. An analysis of variance performed on the mean difference scores (mean rate during No. 9 minus mean rate during No. 10) for Trials T2–6 shows the Numbers effect to be significant at better than the .001 level. The nonsignificant Numbers × Instructions interaction indicates that the magnitude of the deceleration was as great for those *S*s required to control their rate of breathing as for those *S*s allowed to breathe normally. As in the previous studies, the amplitude of both acceleration and deceleration effects was as great for *S*s who were merely told to expect shock as it was for those who actually received it.

Analysis of the respiration records indicates that the breathing of *S*s in Groups C_0 and C_1 was synchronous with the control lights throughout the experiment. For those *S*s in Groups N_0 and N_1 there was no increase in breathing rate during the shock-anticipation trials. The mean rate during Trials B8–10 and T1–6 was 18 cycles per minute during presentation of the numbers. Moreover, *S*s

TABLE 1
ANALYSIS OF VARIANCE OF ACCELERATION AND DECELERATION EFFECTS

Acceleration Effect			Deceleration Effect		
Source	*df*	*F*	Source	*df*	*F*
Between			Between		
Instructions (I)	1	2.83	I	1	3.97
Stimulation (S)	1	2.27	S	1	2.89
I × S	1	0.00	I × S	1	0.08
Error (*MS*)	36	(26597.64)	Error (*MS*)	36	(64250.67)
Within			Within		
Trials (T)	1	16.78***	Numbers (N)	1	36.00***
T × I	1	0.42	N × I	1	0.01
T × S	1	0.06	N × S	1	0.35
T × I × S	1	0.02	N × I × S	1	4.05
Error (*MS*)	36	(8062.57)	Error (*MS*)	36	(1262.71)

*** $p < .001$.

in Groups N_0 and N_1 did not show an increase in time exhaling during No. 10 on the shock-anticipation trials. During the presentation of No. 10 the percentage of time spent exhaling was 57.5 for Trials B8–10 and 56.0 for Trials T1–6. It appears, therefore, that the cardiac changes which have been reported under the present conditions are not attributable to differential respiratory activity.

REFERENCES

DEANE, G. E. Human heart rate responses during experimentally induced anxiety. *J. exp. Psychol.*, 1961, **61**, 489–493.

JENKS, R. S., & DEANE, G. E. Human heart rate responses during experimentally induced anxiety: A follow-up. *J. exp. Psychol.*, 1963, **65**, 109–112.

WESTCOTT, M. R., & HUTTENLOCHER, J. Cardiac conditioning: The effects and implications of controlled and uncontrolled respiration. *J. exp. Psychol.*, 1961, **61**, 353–359.

(Received March 8, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 195–197

VERBAL DISCRIMINATION PERFORMANCE UNDER DIFFERENT VERBAL REINFORCEMENT COMBINATIONS

JANET TAYLOR SPENCE

Veterans Administration Hospital, Iowa City

Performance on a 2-alternative verbal discrimination task was found to be significantly better ($p < .01$) under a Wrong-nothing and a Right-Wrong verbal reinforcement combination than under a Right-nothing combination. Further analyses suggested that these results were due to "nothing" acquiring less reinforcement (information) value when combined with Right than when combined with Wrong.

In investigations utilizing conceptual tasks involving two response alternatives, Buss and his associates (e.g., Buss, Braden, Orgel, & Buss, 1956) reported that performance was poorer under a Right-nothing (Rn) verbal reinforcement combination than under either Wrong-nothing (Wn) or Right-Wrong (RW). In explaining the results of these studies, in all of which *Ss* had been given no prior explanation of the reinforcement procedures, Buss et al. postulated that (*a*) "nothing" is a nonreinforcer, and (*b*) "Wrong" is a stronger negative reinforcer than "Right" is a positive reinforcer. These conclusions appear to imply that in a two-alternative situation, the probability of a *correct* response occurring on a subsequent trial should be greater when the previous response has been *incorrect* and followed by Wrong than when it has been *correct* and followed by Right. However, several two-alternative studies (Buchwald, 1962; Spence, Lair, & Goodstein, in press) have found that it is Right rather than Wrong that produces more correct responses. Further, Buchwald (1959) has shown that nothing in a verbal reinforcement combination acquires reinforcing properties, raising the possibility that the empirical superiority of Wn over Rn was due to the rate at which nothing gains reinforcement value in the two procedures, rather than to the reinforcement values of Right vs. Wrong. Variables associated with conceptual, as opposed, e.g., to rote learning tasks could also play a role. The present study investigated these possibilities.

Method.—The *Ss* were 60 male medical patients from the Veterans Administration Hospital, Iowa City, with no history of neurological or psychiatric problems. A verbal discrimination task (16 pairs of familiar five-letter words) was presented to *Ss* by means of a projector using a film strip in which each pair of words was followed by a blank frame at a 4:2 sec. exposure rate. The *Ss* were given 16 presentations of the list in four serial orders, the right-left position of the members of each pair being counterbalanced across orders.

Instructions stated that *S* was to learn to call out the "correct" member of each pair and that "it will become clear to you as we go along which word in each pair is correct." On the first (guessing) trial prearranged schedules were used so that half of *S*'s initial choices determined the "correct" word and half the "incorrect." The *Ss* were given information by *E* in one of three ways: (*a*) "Right" after correct choices and nothing after incorrect ones (Rn), (*b*) "Wrong" after incorrect choices and nothing after correct (Wn), or (*c*) "Right" or "Wrong" after each choice (RW).

Results and discussion.—The mean percentages of correct responses per trial for each group are plotted in Fig. 1. An analysis of variance of the total number of correct responses per *S* yielded a significant between-groups term, $F(2, 57) = 5.91$, $p < .01$; *t* tests indicated that the performance of the Rn *Ss* was significantly poorer than that of either the RW or Wn groups ($p < .01$) but that there was no significant difference between

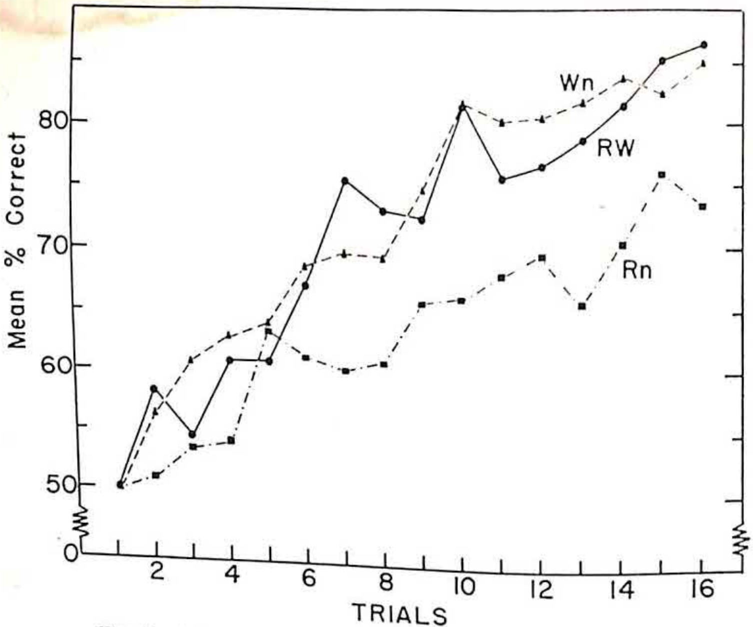


FIG. 1. Mean percent correct per trial for the three groups.

the means of the latter ($t < 1.00$). Thus the results parallel those of prior studies employing conceptual tasks.

In a further analysis the number of correct responses on each trial was determined for each *S* and then the percentage of these responses that continued to be correct on the subsequent trial. Similar percentages for items that had been *incorrect* on the previous trial were also found. The means of these successive percentages on Trials 2–8 for the RW group and for Trials 2–10 for the Rn and Wn groups are plotted in Fig. 2. (Comparisons on later trials were complicated by the presence of *S*s giving 0 correct to 0 previously incorrect responses.) The curves for percent correct following previously correct may be seen to be similar for the three groups, implying that nothing in the Wn group was equally effective as Right in the RW and Rn groups. The curves for previously incorrect items, however, suggest that nothing produced less change to correct responses in the Rn *S*s than Wrong in the RW and Wn groups. Since, except for Trial 2, the percentages were based on different numbers of items within

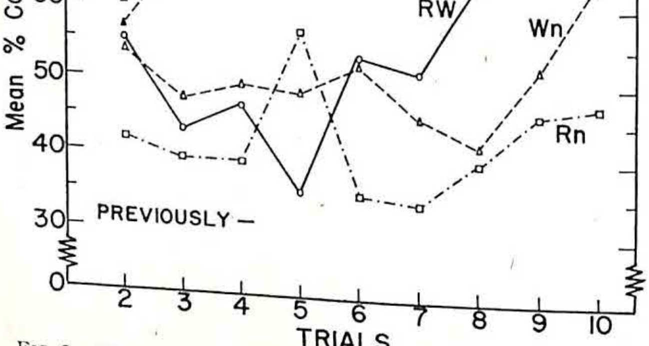


FIG. 2. Mean percent correct on a given trial for items that had been correct (+) and incorrect (−) on the previous trial.

and between groups, only the Trial 2 data were analyzed statistically. A 2 × 2 analysis of variance for correlated measures of number correct following initially correct and incorrect choices in the Wn and Rn groups yielded significant *F*s ($p < .01$) for Groups (due to the superiority of the Wn *S*s on the initially incorrect items) and for the Groups × Measures interaction. A *t* comparing mean number correct following initially incorrect for the combined RW and Wn groups vs. the Rn *S*s was also significant ($p < .01$), but a similar *t* for the combined RW and Rn groups vs. the Wn *S*s on initially correct pairs was not ($p > .05$), again indicating that nothing had less effect in producing correct responses on the Rn *S*s than Wrong on the other groups but an equivalent effect on the Wn *S*s to Right.

In a prior verbal discrimination study (Spence et al., in press) it was found that with *S*s who had been previously instructed about the information value of nothing there were no significant performance differences between Rn and Wn groups and that nothing in each combination had the same effect as an overt response from *E*. These data, in conjunction with the results of the present study, suggest that the performance differences obtained in prior investigations employing uninformed *S*s were due to nothing acquiring reinforcement (information) value less rapidly or to a lesser degree when combined with Right than when combined with Wrong.

REFERENCES

Buss, A. H., Braden, W., Orgel, A., & Buss, E. H. Acquisition and extinction with different verbal reinforcement combinations. *J. exp. Psychol.*, 1956, **52**, 288–295.

Buchwald, A. M. Experimental alterations in the effectiveness of verbal reinforcement combination. *J. exp. Psychol.*, 1959, **57**, 351–361.

Buchwald, A. M. Variations in the apparent effects of "Right" and "Wrong" on subsequent behavior. *J. verbal Learn. verbal Behav.*, 1962, **1**, 71–78.

Spence, J. T., Lair, C. V., & Goodstein, L. D. Effects of different feedback conditions on verbal discrimination learning in schizophrenic and nonpsychiatric subjects. *J. verbal Learn. verbal Behav.*, in press.

(Received March 14, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 2, 197–198

MONETARY INCENTIVES AND VIGILANCE[1]

BRUCE O. BERGUM AND DONALD J. LEHR

United States Army Air Defense Human Research Unit, Fort Bliss, Texas

A visual vigilance experiment was performed in which (a) the effects of monetary incentives, and (b) the effects of removal of these incentives were tested. 20 experimental and 20 control Ss were tested in 2 sessions of 60 and 90 min. each. The experimental group received 20¢ for every signal correctly detected and had 20¢ deducted for every signal missed in the 1st session, but received no reward in the 2nd session. The control group was never rewarded. The rewarded group performed better ($p < .025$) than the controls in the 1st period of the 1st session, and poorer in the final period ($p < .025$) of the 2nd session. These effects were interpreted as resulting from experimentally induced changes in the motivational level of the reward group.

According to Malmo (1959), the level of activation of the ascending reticular activation system is a product of both internally and externally generated stimulation. Thus it would appear that if activation level is the underlying mechanism determining vigilance performance, as has been suggested by Deese (1955) and others, experimental manipulations of either source of stimulation should yield corresponding changes in vigilance performance.

Most of the research in vigilance has involved changes in the level of external stimulation (Bergum & Lehr, 1963; Jenkins, 1958; McCormack, 1958; Webb & Wherry, 1960), while in contrast there is little evidence relating directly to the effects of deliberate motivational manipulations on vigilance performance. The purpose of the present study was to attempt to modify such performance through the direct manipulation of motivation by means of monetary incentives, and to determine the extent of transfer of these effects to a second session in which the monetary incentives were withdrawn.

Method.—A total of 40 trainees from the Army Training Center, Fort Bliss, between the ages of 18 and 26 yr. were employed in two groups. The apparatus was identical to that

---

[1] The research reported in this study was performed for the Human Resources Research Office of George Washington University under contract to the Department of the Army. The opinions expressed in this report are solely those of the authors and do not necessarily reflect those of the sponsoring agency.

TABLE 1

Percent Correct Detections for Both Groups in Both Sessions

Group	Session 1				Session 2			
	Time Periods				Time Periods			
	1	2	3	Average	1	2	3	Average
Incentive	98	74	66	79	84	61	59	68
Control	84	75	68	75	85	75	85	82
Average	91	75	67	77	85	68	72	75

employed in earlier studies (Bergum & Lehr, 1963) and included four test booths, a circular light panel, and an intercom network. The *S* was seated at a small table facing the display which was mounted at eye level on one wall. The 20 lights mounted on the periphery of the circular display illuminated in sequence at a rate of 12 rpm and a signal consisted of the failure of a lamp to illuminate in its normal sequence. The signal rate was 12 signals per hour. Single response push buttons were located in each booth and both signals and responses were automatically recorded on paper tape.

All *S*s were initially given a 20-min. pretest on the apparatus, followed by a 10-min. rest and instruction period in the booths. This was followed by two major monitoring sessions, separated by a 20-min. rest period. In the first session *S*s monitored continuously for a period of 60 min. The 20 control *S*s (Group C) received no special treatment during this session, but the 20 experimental *S*s (Group I) were told that at the end of this session they would receive 20¢ for each signal they had correctly detected, but that 20¢ would be deducted from their totals for each failure to detect a signal. False responses were neither rewarded nor penalized. In the second major session the experimental *S*s were informed that they would no longer be rewarded for their performances and all *S*s monitored continuously under identical conditions for a period of 90 min.

Results.—A comparison of the two groups in terms of pretest performance indicated very similar means and no statistical tests were performed. The mean percent correct detections for the two groups at each of three time periods during each major session are presented in Table 1.

In terms of the first (rewarded) session, examination of Table 1 shows a decrement in performance for both groups with Group I demonstrating a generally higher level of performance than Group C (98% vs. 84%) in the first third of the session. An analysis of variance for this session indicated a significant overall decline in performance over the three time periods, $F\,(2, 76) = 12.15$, $p < .01$. In addition, a significant interaction between time periods and conditions was demonstrated, $F\,(2, 76) = 6.04$, $p < .01$. The t tests indicated that Group I performed significantly better than Group C in the first time period and no better in the last two time periods, $t\,(38) = 2.37$, $p < .025$.

In terms of the second (no-incentive) session, Table 1 indicates a decrement in performance for both groups with Group I demonstrating a generally lower level of performance than Group C in the last two time periods of the session. An analysis of variance of these data indicated a significant overall performance decrement, $F\,(2, 76) = 7.35$, $p < .01$, and a significant interaction between time periods and conditions, $F\,(2, 76) = 4.50$, $p < .05$. Performance of the incentive group was at the same level as that of the control group in the first period, but was significantly poorer in the last period, $t\,(38) = 2.25$, $p < .025$.

Discussion.—The results suggest that the facilitating effect of monetary incentives on vigilance performance is short-lived at best and that, when withdrawn, incentives can in fact be significantly detrimental to performance. The fact that facilitation *did* occur is in line with what would be predicted from the activation model, as is the shift in the general level of performance when incentives were withdrawn. Thus the notion that changes in internal levels of stimulation as well as in external levels of stimulation will lead to concomitant shifts in vigilance performance was substantiated. However, the activation model appears to be too general to have predicted either the loss of effectiveness of the incentive by the end of the first session or the magnitude of the depressing effect on performance resulting from withdrawal of the incentive.

REFERENCES

Bergum, B., & Lehr, D. J. Effects of authoritarianism on vigilance performance. *J. appl. Psychol.*, 1963, **47**, 75–77.

Deese, J. Some problems in the theory of vigilance. *Psychol. Rev.*, 1955, **62**, 359–368.

Jenkins, H. M. The effects of signal rate on performance in visual monitoring. *Amer. J. Psychol.*, 1958, **71**, 647–661.

Malmo, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

McCormack, P. D. Performance in a vigilance task as a function of inter-stimulus interval and interpolated rest. *Canad. J. Psychol.*, 1958, **12**, 242–246.

Webb, W. B., & Wherry, R. J., Jr. Vigilance in prolonged and repeated sessions. *Percept. mot. Skills*, 1960, **10**, 111–114.

(Received March 2, 1963)

Journal of
Experimental Psychology

VOL. 67, NO. 3 MARCH 1964

COMPOSITION OF ASSOCIATIVE CLUSTERS[1]

HOWARD R. POLLIO[2]

University of Michigan

A group of 24 Ss produced associates continuously for 4 min. to each of 4 stimulus words. Although a negative exponential function described the relationship between cumulative association production and time permitted for association quite well for group data, examination of individual association rates showed periods of rapid response bursting intermingled within periods of slower response. An examination of the specific associates contained in these fast sequences showed them to be more highly interconnected in terms of associative overlap, and more highly congruent in terms of connotative meaning than the associates contained in slow sequences. Associative overlap was measured by a matrix analysis of the total number of potential interconnections existing among the words in the fast and slow sequences, while connotative congruence was measured through distance scores computed on the semantic-differential ratings of these words. The results indicated that alternations in overt association rate may be attributed to the existence of associative clusters which are composed of a group of strongly interconnected words, each of which evokes essentially similar meaning responses.

Normative word-association data indicate that there are regularities to the specific associates evoked by a stimulus word. This stability of association tends to indicate that there are enduring patterns of relationship among the constituent words in an associative hierarchy. Recently Deese (1962) has attempted to describe this type of associative structure in terms of the pattern of interitem association present in a group of words, all of which were evoked by the same stimulus word. Evidence for the behavioral effects of associative structure on free recall is to be found in the earlier studies of associative clustering (Bousfield, 1953; Jenkins & Russell, 1952) and in the more recent work of Deese (1959a, 1959b) and of Rothkopf and Coke (1961). These studies have shown that recall may be affected by a simple associative connection between two words as well as by the total pattern of interitem association present among all words in a set of words. Pollio (1963a) has also shown that the latter type of structure predicts the order of emission of words contained in the associative hierarchies aroused by six different stimulus words.

---

[1] This paper is based on a dissertation submitted to the Department of Psychology of the University of Michigan in partial fulfillment of the requirements of the PhD degree and was supported by a United States Public Health Service predoctoral research fellowship. Appreciation is due to A. W. Melton under whose direction this dissertation was conducted, and to Helen Peak, R. W. Gerard, and W. Norman who served on the doctoral committee.

[2] Now at the University of Tennessee.

Interitem associative connections is, however, not the only factor affecting the characteristics of evoked word associates. In one recent investigation (Pollio, 1963b) it was shown that the semantic characteristics conditioned to linguistic stimuli significantly affect the semantic characteristics of associates evoked by these stimuli, in terms of the semantic-differential (SD) ratings of both. Similarly, Pollio (1962) and Staats and Staats (1959) have shown that a word and its associates tend to have similar SD ratings across the three major semantic dimensions described by Osgood, Suci, and Tannenbaum (1957).

If we assume that the speed of overt association is related to the extent of shared structure common to two or more words, then a group of words which has both high interitem associative strength, as well as meaningful congruence, should tend to be produced rapidly in the course of overt association; while words not sharing these properties so strongly should be produced more slowly. A group of words which is both strongly interconnected, and meaningfully congruent, may be called an associative cluster. The constituent words in such a cluster should be emitted rapidly in the course of overt association, while words not contained within a cluster should be emitted more slowly. Early work by Bousfield and Sedgewick (1944) did in fact show that the course of overt continuous association does proceed smoothly; rather individual *S*s show periods of rapid associative responding intermingled within periods of slower response. Response bursting appears irregularly throughout the records of individual *S*s and words contained in these bursts seem to be intuitively related, i.e., to be part of an associative cluster.

The specific purpose of the present study is to determine if variations in the rate of continuous word association can be used to indicate the existence and characteristics of associative clusters. In order to secure data appropriate for this task it is necessary to permit an individual *S* to associate continuously for a given time interval. An examination of each *S*'s individual association rate should reveal extremely rapid associative responding intermingled with periods of slower response. It is assumed that an analysis of these bursts will provide evidence for the hypothesis that an associative cluster is produced rapidly and is composed of a group of strongly interconnected words, each of which evokes essentially similar connotative meaning responses.

METHOD

The overall plan of the present experiment required three separate sessions with each *S*. In the first session *S* associated continuously to four different stimulus words for 4 min. each. An examination of the total series of words produced was made, and three different types of associative sequences were chosen; one in which interword associative response time (RT) was fast, one in which interitem RT was of medium speed, and one in which interitem RT was slow. After all of the words comprising these sequences were selected, half of the *S*s gave eight associations to each word, and then rated each of these words on the following semantic-differential scales: pleasant-unpleasant, soft-hard, and active-passive. The other half of the *S*s first rated the words selected, and then produced associations. Sessions 1, 2, and 3 required 6 wk. A 4-wk. interval occurred between Sessions 1 and 2, and 1 wk. occurred between Sessions 2 and 3. The *Ss* were tested individually in Session 1 and in groups for Sessions 2 and 3.

Subjects.—Twenty-four male undergraduate students participated in this experiment, and each was paid for his participation.

Stimuli.—Four nouns were chosen as stimuli for this experiment: TROUBLE, THIEF, JUSTICE, and HOUSE. These words were chosen on the basis of four criteria: (*a*) the probability of their primary associate; (*b*) their evaluative semantic-differential (SD) rating; (*c*) the evaluative SD rating of their primary associate; (*d*) their frequency of occurrence in the English language. Table 1 presents these stimulus words and their characteristics. Here it may be seen that two of the four stimulus words are connotatively positive and that the other two are connotatively negative. For all four words the probability of their primary associate does not exceed .29. The primary associate to each of these stimulus words is connotatively similar to its own stimulus word. Two of the words may be considered abstract (JUSTICE and TROUBLE), while two (THIEF and HOUSE) may be considered concrete.

All 24 possible orders of these four stimulus words were made. Order of association effects was counterbalanced by having each of the 24 *Ss* associate to each of the four words in one of these 24 orders.

Apparatus and procedure.—The apparatus consisted of a Wollensak tape recorder and microphone; a two-channel Brush pen recorder so arranged as to permit a vocal response to actuate it, a Kodak timer, a Kodak Carousel projector, and an *E*-controlled response key. All of this apparatus was mounted in a large wooden console. During Session 1 *S* sat on one side of the console with only a 9 × 9 in. opaque glass projection surface facing him. Each of the four stimulus words was presented one at a time to *S* for 4 min., during which time *S* was required to "give associations as they are suggested to you."

All of *S*'s associations were recorded on the tape recorder. After Session 1, the latency and duration of each associative response were recorded on one channel of the Brush recorder through direct activation by *S*'s voice on the magnetic tape, and each distinctive word response was recorded on a second channel by *E*. After *S* left the experimental room, two judges reran the tape and transcribed the sequential order of associations produced by each *S*.

The interword associative response latency between each associative response was then measured from the Brush tape to the nearest millimeter. It was from these data that individual cumulative association graphs and individual interword association latency distributions were constructed for each *S*. Group data were also derived from this analysis.

At the conclusion of these preliminary analyses, interword latency distributions for each *S*, in response to each of the four stimulus words separately, were examined. In order to define fast and slow latencies unequivocally, each *S*'s interresponse latency distribution for each stimulus word was divided into quarters. An examination of the total series of words produced by each *S* in response to each of the four stimulus words was then made in order to determine the longest sequence of words, in which each of the latencies separating individual words within the sequence fell in the slowest quarter of the distribution of latencies for each *S*, for each stimulus word separately. These words were taken to define the length of, and specific words in, a slow sequence. Using the length of this slow sequence as a guide, a second set of words, equal in length, was selected. In

TABLE 1

STIMULUS WORDS AND THEIR SEMANTIC CHARACTERISTICS

Stimulus[a] Word	Evaluative[b] SD Rating	Frequence of[c] Occurrence	Prob. of[a] Primary Assoc.	Primary[a] Assoc.	Evaluative SD[b] of Primary Assoc.
TROUBLE	6.10	AA	.09	BAD	6.87
THIEF	6.37	28	.29	STEAL	6.67
JUSTICE	1.67	A	.25	PEACE	1.30
HOUSE	1.77	AA	.25	HOME	1.47

[a] Obtained from Russell and Jenkins (1954) norms.
[b] Obtained from Jenkins, Russell, and Suci (1958) norms.
[c] Obtained from Thorndike-Lorge (1944) word count.

this sequence each of the interword latencies fell within the fastest quarter of the total distribution of latencies. This second set of words defined a fast sequence. All comparisons to be reported in subsequent sections are between fast and slow sequences defined in this way.

In order to clarify the functional relationship between sequence-evocation speed and semantic distance; and between sequence-evocation speed and interword associative connectivity, it was decided to define a medium-sequence evocation speed. Medium evocation speed was defined residually; that is, those latencies falling within the second and third quarters of the distribution of latencies for each *S* in response to each stimulus word, defined medium sequence speed. Since four medium-speed sequences would result in an excessive number of stimuli to be judged in Sessions 2 and 3, one medium-speed sequence was chosen for each *S*. The specific sequence chosen was selected by a random procedure. This procedure resulted in seven medium-speed sequences in response to the stimulus word HOUSE, five to THIEF, and six to JUSTICE and TROUBLE.

After specific words were selected as representative of fast, medium, and slow sequences, association (A), and semantic-differential (SD) booklets were constructed for Sessions 2 and 3. These booklets, contained both the specific words already selected, as well as the four initial stimulus words. The A (association) booklets had each of the words selected for each *S* at the top of a separate sheet of paper. On each of these sheets of paper, the word was typed eight times in two columns of four. (Eight associates were required since the longest sequence of associates selected contained eight words.) A blank space appeared to the right of each of these eight repetitions of the word. The order of words in each of the A booklets was randomly determined, with the provision that not more than two words from the same sequence follow each other in any booklet. Semantic-differential (SD) booklets were constructed according to a similar procedure.

The instructions used in the association task were similar to those used by Noble (1952) in obtaining *m* values for words. Basically, each *S* was asked "to write down eight other words which the key word brings to mind. . . . A good way to do this is to repeat each key word over to yourself as you write." The instructions used in obtaining SD ratings were similar to those used by Jenkins, Russell, and Suci (1958), and asked *Ss*

> to rate each of the key words on the basis of what they mean to you. Place a check mark on each of the SD scales wherever you feel the word should be rated. . . . Rate according to your first impression of the words. Don't hesitate to use the extreme ends of the scales. . . .

The *Ss* then were given a few examples of how these ratings were to be made in order to provide them with anchors for their ratings.

Method for measuring total associative cohesiveness.—In a previous paper (Pollio, 1963a) we suggested that the interword associative relationships existing in a set of words could be described in terms of the $n \times n$ matrix called a *W* matrix, in which the same *n* words appear as both row and column headings. When any word heads a row it is considered as a stimulus and when it heads a column it is considered as a response. For each column word produced as an associate by any row word, a 1 is entered into the appropriate cell in the matrix. If, for example, in an independent association task the word heading Row 3 produced the word heading Column 2 as an associative response, a 1 is entered in Cell w_{32}. After all associative relationships in a set of words have been recorded in this way, an index of cohesiveness (*C*) may be calculated in terms of the sum of the *W* matrix; i.e., $C = \Sigma_i \Sigma_j w_{ij}$, where w_{ij} is a general entry in *W*. The *C* values denote the total number of associative interconnections present in a group of words. If *C* is large we may assume the set to be highly cohesive.

As Luce (1950; Luce & Perry, 1949) has shown for social groups (using essentially the same type of matrix representation) raising Matrix *W* to its various powers indicates the total number of potential multistep interconnections among the initial matrix elements. That is, if Word A associates Word B, but not Word C, and Word B associates Word C, then A has a potential 2-step interconnection with C; i.e., A → B → C. This potential interconnection between A and C is indicated by a 1 in the appropriate cell of Matrix W^2, where W^2 is the square of the initial *W* matrix. In a similar way, more remote potential interconnections are indicated by appropriate entries in the higher power matrices of Matrix *W*. Using binary addition rules (i.e., $0 + 0 = 0$; $1 + 0 = 1$ and $1 + 1 = 1$) in raising *W* to its various powers results in a series of *W* matrices each having 1 and 0 as entries. In order to assess the total number of potential interconnections present in an original set of *n* words, we need only consider Matrices W^1, $W^2 \cdots W^n$, where *n* is the size of the initial matrix W^1. Using binary addition rules Matrix W^{n+1} has exactly the same row and column entries as Matrix W^1. Performing a cellwise summation of Matrices W^1 through W^n results in an A matrix consisting of 1 and 0's. By analogy to the cohesiveness index derived from the original W matrix, it is possible to define a measure of total cohesiveness, MC, for the A matrix in the following way: $MC = \Sigma_i \Sigma_k a_{ij}$ where a_{ij} is a general entry in Matrix A. MC denotes the total number of *potential* interconnections existing among all of the words contained in the original set of words. Large MC values indicate a strong degree of associative cohesiveness in a set of words.

Since not all W matrices (and by implication, all A matrices) are necessarily of equal size, these differences must be taken into account if we are to compare the relative amount of associative cohesiveness existing in two sets of words. Probably the easiest way in which to take differences in initial size into account is to compare the number of words potentially interconnected in a group of words (MC) with the total number of interconnections possible for that group (n^2). A measure of relative cohesiveness, MC_R, may then be defined as follows: $MC_R = \Sigma_i \Sigma_j a_{ij}/n^2$, where a_{ij} is defined as before and where n^2 is the total number of interconnections possible. Since MC_R is a proportion, its value varies between .00 and 1.00. In a highly cohesive set of words, the value of MC_R tends towards 1.00, whereas in a weakly cohesive set it tends toward .00. The value of MC_R is then an appropriate index of the degree of associative cohesiveness present in a group of words.

Method of measuring meaningful congruence.—Since word meaning has been operationalized by the use of SD ratings, the measure $D_{il} = \sqrt{\Sigma d_{il}^2}$ is an appropriate index of the distance between any two words (Osgood et al., 1957). If a set of words has n elements in it, there are $n(n-1)/2$ total distance comparisons possible. Since not all of the word sets which we will consider are of equal size, it is necessary to take these differences into account. Let us, therefore, define a measure of mean intracluster distance $\bar{D}$, which represents the average semantic distance among all elements in a set of words, such that:

$$\bar{D}_i = \frac{D_i}{\frac{n(n-1)}{2}}$$

where D_i is the total distance involving all pairs of words, and $n(n-1)/2$ is the total number of comparisons. The measure $\bar{D}$, denotes the average intraset distance for any set of words in semantic-differential space. Large values of $\bar{D}$ indicate great dissimilarity in connotative meaning among words contained in a sequence of words, while small $\bar{D}$ values indicate meaningful congruence.

RESULTS

Characteristics of continuous association.—At the conclusion of Session 1, 96 individual interword association latency distributions were constructed. Individual interword latency distributions were skewed toward the longer latencies for most *S*s.

From these data, a mean cumulative association curve for each stimulus word for all 24 *S*s was constructed. Figure 1 presents these group cumulative association curves. It is to be noted that the form of each of these curves is similar to that reported by Bousfield and Sedgewick (1944). Their equation, $N = C(1 - e^{-mt})$ where N is the number of associates produced, m is a constant indicating the rate at which associates are produced, C is a constant representing the asymptotic number of associates possible, and t is the amount of time given for continuous association, was found to describe these data extremely well. An evaluation of specific values for m and C constants for each of the four curves was made, and no consistent relations were evident between these two constants for the four stimulus words used in the present experiment. An examination of the distribution of C values, however, did indicate that for these four words the two positively evaluated nouns had more associates available to them than the two negatively evaluated ones.

Classifying these four stimulus words both in terms of evaluative connotation (pleasant vs. unpleasant) and in terms of whether they are abstract (JUSTICE and TROUBLE) or concrete (HOUSE and THIEF), results in a 2 × 2 analysis of variance model.

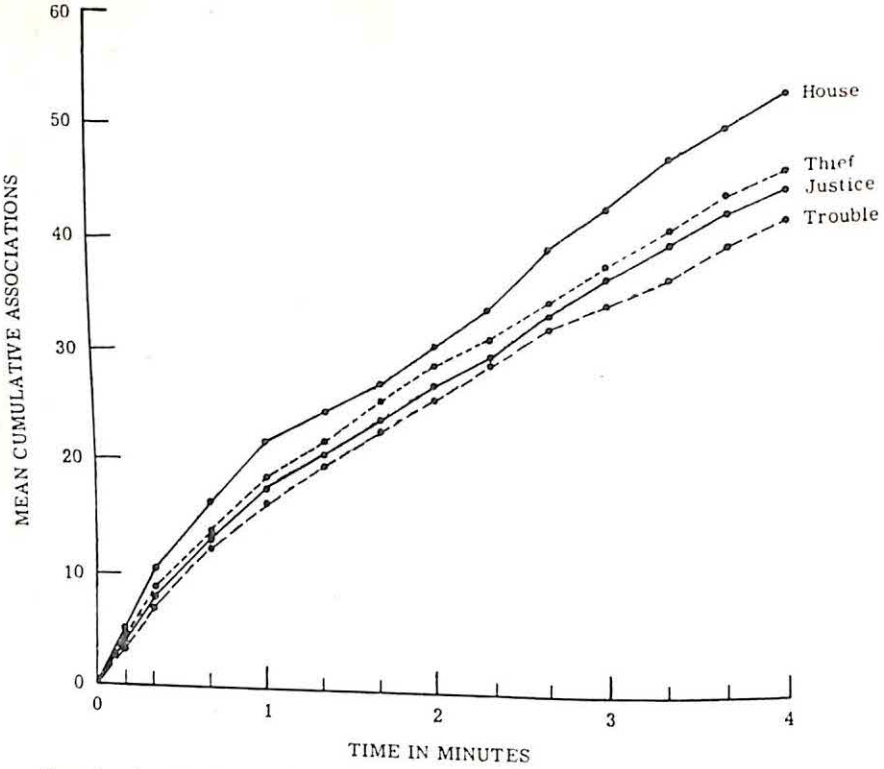


FIG. 1. Mean cumulative continuous association production as a function of association time and initial stimulus word.

The results of a 2 × 2 analysis of variance, involving repeated measurements over *S*s, performed on the terminal number of associations produced by each stimulus word, showed for these four stimulus words, both connotation, $F(1, 23) = 8.03, p < .01$, and abstractness, $F(1, 23) = 12.05$, $p < .01$, significantly affected the total number of associates produced.

Intracluster associative cohesiveness.—From the data provided by the individual interword association latency distributions the specific words composing the fast-, slow- and where appropriate, medium-speed sequences were selected. After all of the data from Sessions 2 and 3 were collected these sequences were represented by *W* matrices, separately constructed

TABLE 2

MEANS AND *SD*s OF RELATIVE INTRASEQUENCE COHESIVENESS (MC_R) AS A FUNCTION OF SEQUENCE SPEED AND STIMULUS WORD

Sequence Speed	Stimulus Word							
	HOUSE		JUSTICE		THIEF		TROUBLE	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Fast	.222	.161	.308	.294	.276	.266	.185	.236
Slow	.046	.088	.029	.056	.024	.045	.022	.037

for the associates of each S in response to each stimulus word. Each of these W matrices was multiplied out to its own nth power, and A matrices then constructed from these results. MC_R measures were computed from each S's total A matrix.

Table 2 presents the means and *SD*s of the intrasequence MC_R's for all *S*s as a function of both the stimulus word evoking the sequence and the speed with which the sequence was produced. In all cases, as predicated, fast sequences were found to be more cohesive than slow sequences. There are no reversals in this trend. Since for five of the eight entries in Table 2, the *SD*s exceeded the means, MC_R values were transformed by a $\sqrt{X + .5}$ transformation, and then subject to a 4×2 analysis of variance in which one of the independent variables was sequence speed, and the other, stimulus words. The results of this analysis showed that only sequence speed significantly affected the cohesiveness of associative sequences, $F\ (1, 23) = 58.46$, $p < .001$. This indicates that rapidly evoked associative sequences are composed of elements which are more highly interconnected than is true for elements contained within associative sequences which are produced more slowly.

In order to examine the functional relationship between associative cohesiveness and sequence-evocation speed across at least three different speeds, MC_R values for medium-speed sequences were also computed. Table 3 presents the mean MC_R values and their associated *SD*s for the 24 randomly selected fast, medium, and slow sequences. As can be seen, MC_R values are ordered as expected. Fast sequences are maximally cohesive, slow sequences are minimally cohesive, and the cohesiveness of medium-speed sequences falls between these extreme values.

TABLE 3

MEANS AND *SD*s FOR THE MC_R VALUES OF 24 RANDOMLY SELECTED FAST, MEDIUM, AND SLOW SEQUENCES

Sequences Speed	Mean MC_R Proportion	*SD*
Fast	.229	.242
Medium	.083	.097
Slow	.021	.036

An examination of the individual means and *SD*s again indicates a marked skewness in MC_R distributions with a great many .00 MC_R values occurring at all sequence-evocation speeds. Consequently, these data were transformed by a $\sqrt{X + .5}$ transformation. The results of an analysis of variance done on these transformed scores again confirmed the expectation of an increasing relationship between associative-evocation speed and the cohesiveness of an underlying cluster, as the only significant source of variation among MC_R means could be attributed to sequence speeds, $F\ (2, 40) = 11.63$, $p < .001$.

Intracluster semantic distances.—After the SD ratings of all words contained in fast, medium, and slow sequences for each S were collected, a semantic-distance matrix involving all words in a given sequence was constructed. Each of these matrices contained $\frac{n(n-1)}{2}$ entries, where n is the number of words selected for each S. The words are obviously exactly the same words as were used in constructing W matrices for the preceding analyses of intrasequence cohesiveness.

$\bar{D}$ measures were computed between each of the $\frac{n(n-1)}{2}$ pairs contained in each semantic-distance matrix.

TABLE 4

MEANS AND *SD*s OF THE $\bar{D}$ MEANS AS A FUNCTION OF SEQUENCE SPEED AND STIMULUS WORD

Sequence Speed	Stimulus Word							
	HOUSE		JUSTICE		THIEF		TROUBLE	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Fast	3.61	1.10	2.83	1.21	3.62	1.30	3.35	1.44
Slow	4.27	.97	3.81	.96	4.46	1.09	4.26	1.51

Table 4 presents the means and *SD*s of the intrasequence $\bar{D}$ means for all *S*s as a function of both the stimulus word evoking the sequence, and the speed with which the sequence was evoked. As predicted, the mean semantic distance between all words in a fast sequence is less than is the mean distance between all words in a slow sequence. This relationship holds across all four stimulus words, and there are no reversals in this trend.

An analysis of variance performed on these data indicated that differences between Speeds, F (1, 23) = 28.51, $p < .001$, and between Words, F (3, 69) = 4.58, $p < .01$, were all highly significant. As expected, the mean semantic distance was significantly greater for slow sequences than for fast sequences.

In order to demonstrate the generality of an inverse relationship between associative-production speed and intracluster semantic distance, the mean $\bar{D}$ of 24 medium-speed sequences was also examined. As can be seen in Table 5, the order of mean $\bar{D}$ magnitude is inverse to the speed of sequence evocation. This trend is quite clear and consistent with the previous analysis. An analysis of variance was applied to these data, and the only significant source of variance among $\bar{D}$s was produced by differences in the speed of sequence evocation, F (2, 40) = 5.70, $p < .01$. The results of this analysis indicate that there is an inverse relationship between the speed of an associative sequence's evocation and the magnitude of its underlying intracluster semantic distance.

TABLE 5

MEANS AND *SD*s OF THE $\bar{D}$ MEANS OF 24 RANDOMLY SELECTED FAST, MEDIUM, AND SLOW SEQUENCES

Sequence Speed	Mean $\bar{D}$	*SD*
Fast	3.17	1.09
Medium	3.76	1.11
Slow	4.05	1.11

DISCUSSION

The results of the present experiment suggest that overt continuous word association is to be understood in terms of the pattern of relationships which exist among words and "that this arrangement implies the course that activity will take when this structure is activated [Peak, 1958, p. 325]." As we have demonstrated, two significant dimensions to this pattern are to be found in the associative cohesiveness of words and in the connotative meanings of words.

In a recent conference on verbal behavior (Cofer, 1961) there was some dispute as to the exact nature of *the* "representational" mediator. None of the discussants denied the importance of mediational processes, rather their differences arose in describing the specific nature of the mediating event. Bousfield (1961) took the position that only words should be considered as appropriate vehicles for the mediation function while Osgood (1961) argued that conditioned meaning responses were of primary importance. The results of the present experiment suggest that both types of mediational events are involved in determining the rate of emission of word associates in a continuous association task, and that, at least at the present stage in our understanding of the association and mediation processes, nothing is to be gained by reducing one to the other. The major theoretical and empirical task would now seem to be one of discovering under what antecedent conditions selection between these mediators occurs, if it does occur, rather than one of denying the existence of either of these two events.

What are we to make of the obtained differences among stimulus words in terms of the number of associations produced? Although only two words in the positive and negative evaluative categories were used in the present study, the data indicate that our positively evaluated words produced more associates than did the negatively evaluated words. This differential effect is quite consistent with recent results reported by Koen (1962). Together, these results may indicate that the density of elements in the positively evaluated portion of the semantic space is greater than in the negatively evaluated portion.

The mean continuous association curves across groups were quite well described by the equation $N = C(1 - e^{-mt})$. An examination of the individual curves, however, made it apparent that interpretation of the m constant is meaningful only for group data. Very few, if any, of the 96 individual records showed a constant rate of association; rather alternation in association rate was much more common.

With respect to these individual alternations in the production of associations, we have seen that greater intracluster associative cohesiveness describes the associative relations which exist among elements in fast sequences relative to slow sequences. A direct increasing relationship was found between three different sequence speeds and the cohesiveness of their underlying clusters.

Although the maximum value possible for an MC_R proportion is 1.00, the highest mean value obtained in the present experiment was .308. An examination of the *SD* associated with each mean MC_R proportion indicated marked skewness in the MC_R distribution such that many sequences yielded MC_R values of .00. These results indicate that for individual *S*s, even in quite rapidly occurring sequences, the associative interconnections among elements are not completely fixed, and despite the obtained associative stability, tend to vary widely from time to time.

Perhaps an example will prove helpful. A fast sequence aroused by the stimulus word TROUBLE for one of the *S*s in the present experiment contained the following three words in the following order: MORNING, NOON, and NIGHT. When a *W* matrix was established for these words for this *S*, it was found that no one word produced any of the other two as associates in the independent assessment procedure (Session 2). Rather his associations to MORNING were: COLD, BREAKFAST, etc., to NOON: LUNCH, RADIO, etc., and to NIGHT: DARK, FUN, etc. The matrix analyses of this pattern obviously yielded an MC_R value of .00. This value is contrary to an intuitive estimate of the associative cohesiveness of the words contained in this sequence. What these data indicate most clearly is that while stable associative patterns do exist within an individual, the overt production of a specific word association may also depend upon other factors. Prediction of a specific associate, or even of a class of associates, depends, as Howes and Osgood (1954) have demonstrated, upon the linguistic context provided by preceding words. It seems clear that in the presence of the verbal stimulus compound, MORNING, NOON . . . the associate NIGHT is much more probable than in the presence of either of these words singly. We would assume the tendency to produce NIGHT after the verbal compound, MORNING, NOON is the result of a summation of relatively weak associative tendencies to the response NIGHT from each of these words separately. Only under the multiple input pattern provided by the compound, does the associate NIGHT, become more probable.

The present experiment also indicates that there is a different pattern of within-cluster connotative relations for elements contained in rapidly occurring, as compared to more slowly occurring, associative sequences. For all sequences regardless of initial stimulus word, the within-cluster semantic distance is less for rapidly occurring sequences. This inverse relationship between semantic distance and speed of sequence evocation also described the ordering of semantic distances across the three different sequence-evocation speeds. The results of this analysis, taken in conjunction with the significant correlations reported by Staats and Staats (1959) and by Pollio (1962) between the SD rating of a Kent-Rosanoff stimulus word and the SD rating of either its total range of associates, or of its primary associate, respectively, imply a total organization of words in which a word and its associate(s) are connotatively similar, and consequently, proximal in semantic space.

REFERENCES

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, **49**, 229–240.

BOUSFIELD, W. A. The problem of meaning in verbal learning. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961.

BOUSFIELD, W. A., & SEDGEWICK, C. An analysis of the sequences of restricted associative responses. *J. gen. Psychol.*, 1944, **30**, 149–165.

COFER, C. N. (Ed.) *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961.

DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312. (a)

DEESE, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. exp. Psychol.*, 1959, **58**, 17–22. (b)

DEESE, J. On the structure of associative meaning. *Psychol. Rev.*, 1962, **69**, 161–175.

HOWES, D. H., & OSGOOD, C. E. On the combination of associative probabilities in linguistic contexts. *Amer. J. Psychol.*, 1954, **67**, 241–258.

JENKINS, J. J., & RUSSELL, W. A. Associative clustering during recall. *J. abnorm. soc. Psychol.*, 1952, **47**, 818–821.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. An atlas of semantic profiles for 360 words. *Amer. J. Psychol.*, 1958, **71**, 688–699.

KOEN, F. Polarization, *m* and emotionality in words. *J. verbal Learn. verbal Behav.*, 1962, **1**, 183–187.

LUCE, R. D. Connectivity and generalized cliques in sociometric group structure. *Psychometrika*, 1950, **15**, 169–190.

LUCE, R. D., & PERRY, V. A method of matrix analysis of group structure. *Psychometrika*, 1949, **14**, 95–116.

NOBLE, C. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

OSGOOD, C. E. Comments on Professor Bousfield's paper. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PEAK, H. Psychological structure and psychological activity. *Psychol. Rev.*, 1958, **65**, 325–347.

POLLIO, H. R. Word association as a function of semantic structure. Unpublished doctoral dissertation, University of Michigan, 1962.

POLLIO, H. R. A simple matrix analysis of associative structure. *J. verbal Learn. verbal Behav.*, 1963, **2**, 166–169. (a)

POLLIO, H. R. Word association as a function of conditioned meaning. *J. exp. Psychol.*, 1963, **66**, 454–460. (b)

ROTHKOPF, E., & COKE, E. The prediction of free recall from word association measures. *J. exp. Psychol.*, 1961, **62**, 433–438.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8-onr-66216.

STAATS, A. W., & STAATS, C. K. Meaning and *m*: Correlated but separate. *Psychol. Rev.*, 1959, **66**, 136–144.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

(Received May 6, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 209–214

INTERSTIMULUS INTERVAL EFFECTS IN GSR DISCRIMINATION CONDITIONING[1]

RUSSELL A. LOCKHART AND WILLIAM W. GRINGS
University of Southern California

GSR discrimination conditioning of college Ss was studied as a function of CS-UCS interval variation. With a .5-sec. delay only 1 CR occurred on acquisition test trials; with a 5-sec. delay 2 CRs were observed: a 1st response during the delay, and a 2nd response following CS termination (with UCS omitted). Maximum discrimination occurred after only 2 differential reinforcements in both interval conditions; CS-UCS interval variation produced no differences in 1st response discrimination; 2nd response discrimination was superior to 1st response discrimination obtained in either interval group. Results were compared with those obtained from mental defectives. The comparison of 1st response discrimination supported the notion that the mental defective attained "perception through conditioning," while the college S showed conditioning-like behavior "through perception." 2nd response discrimination occurred with equal speed in the mental defective as in the college S.

Although the interstimulus interval is considered a crucial variable in classical conditioning, its effects have been studied primarily with the simple conditioning paradigm. Relatively little literature exists on the role of the CS-UCS interval in the more complex discrimination situation. Recently, however, interest has developed in this area.

Grings, Lockhart, and Dameron (1962) demonstrated that for mentally deficient *S*s, GSR discrimination performance was more rapid with a 5-sec. interstimulus delay than with a .5-sec. delay. The comparison between the two interval conditions was not a simple one, however, for it was found essentially impossible to assess the degree of conditioned responding on a comparable basis. Because of the long latency characteristic of the GSR, the response occurs *after* the CS-UCS pairing in the brief interval situation. The longer interval condition permits responding *prior* to UCS onset. In addition to this "anticipatory" response during the long delay period, a "second" response was found to occur during the normal UCR period on acquisition test trials *when the UCS was withheld.* In contrast, no such double responding occurred in the shorter interval group. Indeed, it was in terms of this second response that GSR discrimination performance was greater under the long delay condition, for no difference was found between anticipatory response discrimination and discrimination obtained under the short delay condition.

Recently, more parametric studies have shown that the CS-UCS interval "function" may be quite different in discrimination than in simple conditioning (where the .5-sec. CS-UCS interval has been reported as optimal). Hartman and Grant (1962) compared eyeblink discrimination performance of four groups conditioned with different CS-UCS intervals (400, 600, 800,

---

[1] Supported by a grant to the second author from the United States Public Health Service (Grant M-3916). The authors are indebted to Tadao Uno for his assistance in collection of data.

and 1,000 msec.). They found degree of discrimination to be an *increasing* function of the interval of delay. Kimmel and Pennypacker (1963) employed four intervals (.25, .50, 1.0, and 2.0 sec.) in conditioning the GSR. Again, they found the degree of discrimination to be an increasing function of the delay between stimulus onsets.

The present study extends these results in terms of comparing GSR discrimination behavior in two groups conditioned under more widely separated CS-UCS interval conditions than previously employed in a college population (i.e., .5 and 5 sec.). Hartman and Grant (1962) and Kimmel and Pennypacker (1963) found that their results were due to a decrease in response to the nonreinforced stimulus in the longer delay rather than an increase in response to the reinforced stimulus. If such an effect continues to operate when the delay is extended to 5 sec., then discrimination performance should be more rapid in the longer than in the shorter delay condition.

METHOD

Subjects.—Forty paid undergraduate volunteers served as *S*s. All were found to have normal hearing as determined by an audiometric examination.

Apparatus.—Two stimuli were used as CSs. A 1,000-cps tone produced by an oscillator operating through an amplifier was attenuated to 5 db. above the sound level of the laboratory. The tone was delivered to *S* through a loudspeaker, located to the left and about 2 ft. away from *S*. A 10 ft-L light was produced by a 2 × 2 ft. light box housing eight incandescent bulbs behind a translucent surface. General laboratory illuminance was 1 ft-c. The light box was placed on a table in front of and about 2 ft. away from *S*.

Electrical stimulation was employed as the UCS. The shock was delivered to the volar surface of the right forearm through 1.5 × 1.3 cm. silver electrodes (spaced 4 cm. from center to center) from a Grass Model S-5 stimulator set for monophasic pulses at 20 cps and 2 msec. duration.

Stimulus durations were controlled by Hunter electronic timers. CS duration was 5 sec. in Groups I and II, with the UCS occurring at CS termination in Group I and .5 sec. after CS onset in Group II. In Group III CS duration was .5 sec. with UCS occurring at CS termination. Thus, the CS-UCS interval was 5 sec. in Group I and .5 sec. in Groups II and III. The UCS duration was .6 sec. for each group. Intertrial intervals were manually varied irregularly among 15, 20, and 25 sec.

The GSR was obtained as a dc resistance change through $\frac{1}{2} \times \frac{5}{8}$ in. silver electrodes bent to the contour of *S*'s first and third fingers of the left hand, coated with electrode paste and taped firmly to the finger tips. Recording of the GSR was accomplished by a modified Darrow-type bridge input to a two-channel Dynograph amplifier-recorder (Offner Type 542).

Design.—The *S*s were assigned to one of the three groups defined previously. Group I consisted of 20 *S*s while Groups II and III consisted of 10 *S*s each.[2] Half the *S*s in each group received tone as the positive (reinforced) CS and light as the negative (nonreinforced) CS. The other half received the opposite condition. The *S*s received 6 presentations of each stimulus during an adaptation series and 20 presentations of each stimulus during the subsequent acquisition series. Only one stimulus received reinforcement during acquisition. For this stimulus, a partial reinforcement schedule was used (to allow for test trials during acquisition) resulting in 12 reinforced trials and 8 nonreinforced trials for the positive CS. The latter trials served as "test" trials and occurred on presentations 3, 5, 8, 10, 13, 15, 18, and 20 of the positive CS for all *S*s. The negative stimulus prior to (but not necessarily adjacent) each test trial was designated as a "control" trial for the purpose of comparison with the "test" trial response.

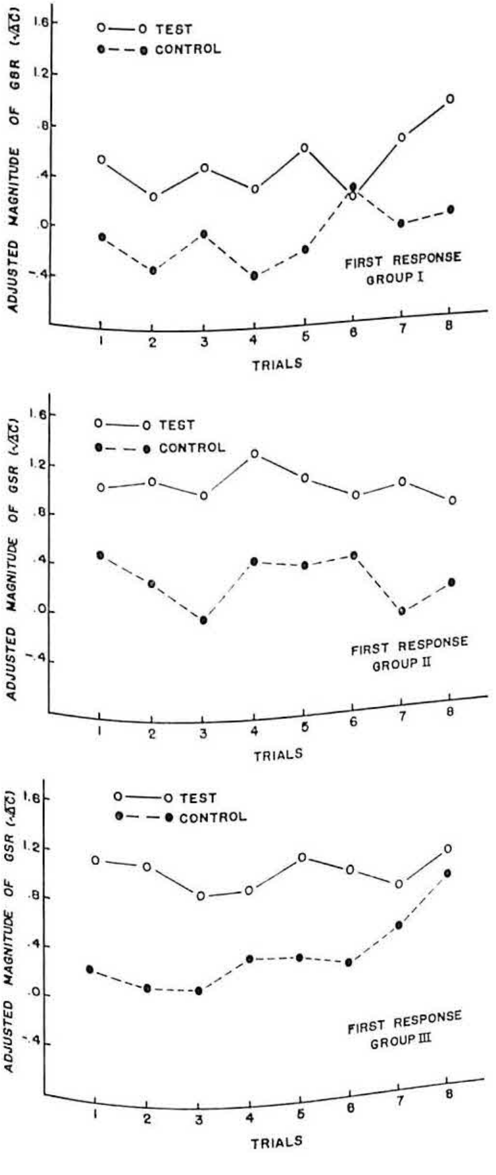
Procedure.—The design and procedure of the present study were equivalent to an earlier study with mental defective *S*s (Grings et al., 1962). However, *S*s in the present study were informed that shock was to be administered and that they would set the intensity of it themselves. This was not possible with the mental defective. No other instructions were given at any time concerning the nature of the stimulus conditions.

---

[2] A long CS was employed in Group II in order to determine the effect of stimulus offset as a possible determiner of the second response observed when the UCS is withheld at CS termination in Group I. In addition, it provides a condition where the stimulus is contiguous with the CR, whereas the CS when very brief terminates prior to the CR.

Results

Responses measured.—Two responses were defined and measured in Groups I and II on each test and control trial. The first was the response during the 5-sec. portion of the CS and will be termed the "first response." The second was the response occurring in the 5-sec. period following CS termination and will be termed the "second response." In Group III, only the response in the 5-sec. period following CS onset was measured since this is the only response obtained under short CS and short CS-UCS interval conditions. All response amplitudes were measured in reference to the base resistance at CS onset and transformed into square root of change-in-conductance units.



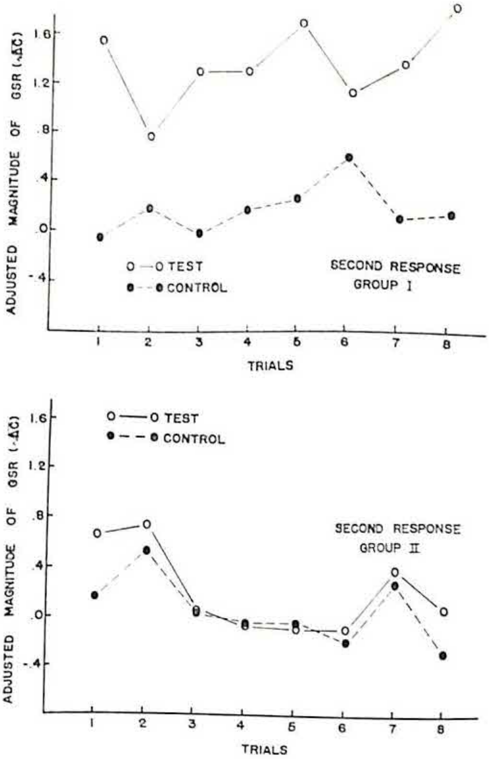


FIG. 1. Magnitude of GSR on acquisition test and control trials for "first" and "second" responses in Group I (5-sec. CS, 5-sec. CS-UCS interval), Group II (5-sec. CS, .5-sec. CS-UCS interval), and Group III (.5-sec. CS, .5-sec. CS-UCS interval). (Responses adjusted for mean level of responding during prior adaptation series.)

Acquisition.—Figure 1 presents the mean magnitude of GSR to positive and negative CSs during acquisition test and control trials. These curves

TABLE 1

MEAN RESPONSE TO TEST AND CONTROL STIMULI AT FIRST POINT OF COMPARISON DURING ACQUISITION; ADJUSTED FOR ADAPTATION LEVEL OF RESPONDING

Group	Response	Test	Control	Difference	t
I	First	.56	−.06	.62	3.23**
I	Second	1.55	−.07	1.62	5.08**
II	First	1.10	.49	.61	2.46*
II	Second	.67	.18	.49	.97
III	First	1.17	.27	.90	2.61*

* $p < .05$.
** $p < .01$.

represent "adjusted" values obtained by subtracting from each GSR amplitude the corresponding mean magnitude of response elicited by the stimulus during the adaptation trials. This procedure was employed as a control for differential responsiveness between the different groups and between stimuli prior to reinforcement within each group. No significant differences were found between the positive and negative stimuli "to be" for any response in any group at any point during the adaptation trials.

Table 1 presents the mean magnitude of adjusted test and control responses at the first point of comparison during acquisition. This point followed two differential reinforcements (i.e., two reinforced and two nonreinforced trials in the order CS+, CS−, CS−, CS+). The most striking feature of the curves in Fig. 1 is that the degree of differential responding was, in general, maximal at the first point of comparison. Table 1 shows quite significant discrimination for both first and second responses in Group I. Less reliable, but significant discrimination was also obtained for the first response in Groups II and III. The second response in Group II failed to show discrimination.[3]

Inspection of Fig. 1 indicates that discrimination performance was relatively constant throughout acquisition. There was no indication that the magnitude of response to the nonreinforced stimulus decreases during acquisition in either the long or short interval groups. Likewise, there appears to be no consistent increase in response magnitude to the positive CS in either interval group.

Although it was not generally possible to analyze responding prior to the first point of comparison during acquisition, it was possible to do so for the first response in Group I. Table 2 shows the magnitude of GSR on the last trial during adaptation and the first two trials during acquisition for both the CS+ and the CS−. The response to CS− failed to increase over the level of response during adaptation. The first trial during acquisition on CS+ was actually a seventh adaptation trial since the UCS did not occur for the first time until CS

---

[3] It should be noted here that significant discrimination performance occurred for the second response in Group II at the first test-control point of comparison when the last response to "test" and "control" stimuli was used in the adjustment procedure (this being the only result which changed as a result of employing the last response rather than the average response during adaptation). The most obvious explanation for such discrimination at a point quite removed from the point of reinforcement would involve backward conditioning, i.e., the UCS terminating prior to stimulus offset (which in this case would serve as the CS). However, this effect rapidly disappeared.

TABLE 2

Magnitude of GSR Elicited by CS+ and CS− during Final Trial of Adaptation Series and First Four Ordinal Trials of the Acquisition Series

Stimulus	Adaptation Trial 6	Acquisition			
		Trial 1	Trial 2	Trial 3	Trial 4
CS+	1.51	1.01	—	—	1.89
CS−	1.24	—	1.26	1.30	—

termination, hence the lower responding (i.e., continuing adaptation) on this trial as compared with Adaptation Trial 6. However, the first response on the second presentation of CS+ was significantly greater than the last presentation of this stimulus during adaptation, and more importantly, greater than the second presentation of CS− during acquisition, *t* (19) = 3.65.

First response discrimination performance in Group II was not different from that obtained in Group III, *t* (18) = .70. Second response discrimination was significantly greater than first response discrimination in Group I, *t* (19) = 3.18. To compare discrimination across CS-UCS interval conditions, the .5-sec. groups were combined. A significant Discrimination × Interval interaction was obtained in terms of the second response of the long group compared to the first response of the short group, *F* (1, 38) = 4.05. When the first responses in both interval groups were compared, the interaction was insignificant, *F* (1, 38) = .24.

DISCUSSION

No differences were found between the .5-sec. and 5-sec. CS-UCS interval conditions when compared on the basis of first response discrimination, although significant discrimination occurred in each group. Second response discrimination in the 5-sec. group was significantly greater than first response discrimination in the .5-sec. group. However, this statement reflects a complicated and possibly illogical comparison of discrimination performance when there is manipulation in the CS-UCS interval parameter since there is no comparable "second response" in the short interval condition. For this reason it may be incorrect to speak of a strict CS-UCS interval "function" in GSR discrimination conditioning.

If the actual response process (GSR) plays any crucial role in the conditioning process, then it would be expected that stimulus conditions allowing GSR responding prior to UCS would lead to different response behavior than when conditions permit no such "anticipatory" responding. However, the results indicate that in terms of the first response occurring to the CS, the CS-UCS interval variation had no effect (i.e., the Discrimination × Interval interaction was insignificant). Apparently for college *S*s the .5-sec. delay is sufficient to permit the development of a discriminatory process which shows itself immediately in terms of differential autonomic behavior. Since the CS-UCS pairing terminates prior to the actual GSR occurrence, the discriminatory process may develop prior to autonomic involvement. It would appear that the actual response system plays little part in the development of the discrimination. It might be hypothesized that the GSR as a response was not directly conditioned to the CS, but more nearly *reflected* a conditioned cognitive or perceptual discrimination. Since no difference was found between the short and long interval groups this conclusion may apply to the longer interval condition as well.

It would appear that the *fact* of differential reinforcement is more important in producing discriminatory behavior than is the temporal relation in which such reinforcement is presented. This statement is not simply an obvious one, for some theories of classical conditioning hold that little or no GSR conditioning occurs when the CS-UCS interval exceeds 1 or 2 sec. (e.g., Moeller, 1954; White & Schlosberg, 1952). If this held true in the discrimination situation, then temporal delay between CS and UCS onsets would be a more determinant variable than the actual differential reinforcement. (It should be pointed out that the stimuli in the present study were qualitatively different—tone and light. This may have facilitated discrimination. Had the stimuli been varied quantitatively—e.g., wavelength or frequency—the CS-UCS interval might have played a more influential role.)

The interstimulus interval parameter was equally ineffective in a mental defective population in terms of first response discrimination (Grings et al., 1962). However, in that population, first response discrimination was found to be an increasing function of number of differential reinforcements, a result not found in the college population. It could be hypothesized that while the GSR behavior of the mental defective represented a conditioned discrimination in the usual sense, GSR behavior of the college *S* represented more nearly a response to an immediately developed perception of stimulus relations. More simply, the mental defective attained "perception through conditioning" (Razran, 1955), while the college *S* showed conditioning-like behavior "through perception."

While such differences between college *S*s were not unexpected, it was surprising to observe that second response discrimination occurred *at the same point* in acquisition in both populations (i.e., after two differential reinforcements). It is conceivable that second response discrimination represents a more basic process in classical conditioning than first interval discrimination. Second response discrimination may develop prior to or be unaffected by verbal-perceptual control, while the first response may be readily modified by verbal-perceptual factors.

Regardless of the interpretation placed upon these data, second response discrimination continues to pose a problem for theories of classical autonomic conditioning.

REFERENCES

GRINGS, W. W., LOCKHART, R. A., & DAMERON, L. E. Conditioning autonomic responses of mentally subnormal individuals. *Psychol. Monogr.*, 1962, **76**(39, Whole No. 558).

HARTMAN, T. F., & GRANT, D. A. Differential eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1962, **64**, 131–136.

KIMMEL, H. D., & PENNYPACKER, H. S. Differential GSR conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1963, **65**, 559–563.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.

RAZRAN, G. Conditioning and perception. *Psychol. Rev.*, 1955, **62**, 88–95.

WHITE, C. T., & SCHLOSBERG, H. Degree of conditioning as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

(Received March 27, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 215–220

EFFECT OF INSTRUCTIONAL SET ON RESPONSES TO COMPLEX SOUNDS[1]

STANLEY J. RULE[2]

University of Washington

90 Ss rated sounds under loudness, noisiness, or annoyance instructions. The sounds were composed of a random noise with discrete tone components. 4 factors of sound were investigated: (a) overall intensity, (b) fundamental tone frequency, (c) fundamental tone intensity, and (d) overtone intensity. An analysis of the interactions between instructions and stimulus factors indicated that instructions set Ss to give different emphasis to stimulus factors. Overall intensity was emphasized most under loudness instructions, followed by noisiness and annoyance instructions. Fundamental tone frequency and fundamental tone intensity were given greatest emphasis under annoyance instructions, followed by noisiness and loudness instructions. No difference was found for overtone intensity.

One fundamental problem in psychoacoustics has been that of relating physical measures of sound to psychological measures. In many cases the operations which define the psychological measures are the instructions given to *S*s in which *S*s are instructed to respond to acoustical stimuli in terms of a subjective experience. However, it is not necessary to assume that a particular subjective state is being tapped by this procedure. The function of the instructions may be that of setting *S*s to respond to particular aspects of the stimulus situation. The purpose of the present study was to determine which aspects of sound are emphasized under three instructional sets: loudness, noisiness, and annoyance.

Research by Laird and Coye (1929) indicated that *S*s give greater emphasis to the frequency of pure tones under annoyance instructions than under loudness instructions. Reese and Kryter (1944) determined equal annoyance and equal loudness contours for bands of noise 250 mels wide. Their findings were similar to those of Laird and Coye, but less pronounced.

Stevens (1956) developed the mark II version of the loudness scale to measure the loudness of sound spectra several octaves wide. A modification of the mark II loudness scale has been made by Kryter (1959, 1960) on the basis of the study previously cited (Reese & Kryter, 1944). The resulting scale was intended to measure *S*s' responses to the noisiness of broad-band noise spectra. By substituting the term noise in the above scale for annoyance, the term originally used in Reese and Kryter's research, Kryter has evidently assumed that the two terms yield equivalent results. However, no data are available concerning possible differences in responses to sounds under noisiness and annoyance instructions.

An additional problem is the lack of adequate research on the effects of discrete tones on judgments of noise spectra. Neither the mark II loudness scale nor the perceived noisiness scale were based on stimuli with discrete tone components. This lack of data may have led to misinterpretation of results. For example, Copeland, Davidson, Hargest, and Robinson (1960) instructed *S*s to equate various jet aircraft sounds for loudness. Their findings indicated a closer relationship of *S*s' loudness judgments with the noisiness scale than with the mark II loudness scale. Robinson and Bowsher (1961) found similar results when *S*s were instructed to equate the loudness of helicopter noise with a jet airplane noise.

---

[1] This report is based on a dissertation submitted by the author in partial fulfillment for the PhD degree in psychology at the University of Washington. Special acknowledgment is due to Sidney S. Culbert, thesis chairman.

[2] Now at the University of Alberta.

These findings if taken at face value would indicate that the perceived noisiness scale is a better measure of loudness than the mark II loudness scale. Since jet airplane noise contains discrete high pitch tone components, the results may have been due to the exclusion of discrete frequency components in construction of the two scales. Discrete frequency components may contribute an increment to *S*s' judgments above that contributed by the overall band intensity. If this assumption is true a closer relationship would be expected between the noisiness scale and loudness judgments of noise which contained high-pitch discrete tone components. Since the noisiness scale gives greater weight to the higher octaves than the mark II loudness scale, the perceived noise scale may have inadvertently given proper weight to the discrete frequency components.

To provide additional information on the effects of discrete tone components on judgments of sound, the stimuli in the present study consisted of a broadband noise which contained a fundamental tone and its first overtone. The stimuli varied along four dimensions: (*a*) overall intensity, (*b*) fundamental tone frequency, (*c*) fundamental tone intensity, and (*d*) overtone intensity. This study was designed to ascertain the relationship of loudness, noisiness, and annoyance instructional sets to each of the four stimulus dimensions.

METHOD

Apparatus.—The apparatus consisted of an Ampex 601-1 tape recorder, six reels of magnetic tape, and five Brush Model BA-206 earphones. The earphones were matched within 2 db. from 50 to 10,000 cps. Recorded on each tape were sounds to be presented to *S*s. The tapes differed only with respect to order in which the sounds appeared. The order of presentation was random.

The auditory stimuli consisted of 36 pairs of sounds. The first sound in each pair was a broadband noise at 76 db. SPL. The broadband noise was produced by a random noise generator and shaped to yield equal energy per octave.

The second sound in each pair was one of 36 complex sounds, consisting of a broadband noise with discrete tone components. The broadband noise had equal energy per octave. The discrete tones were produced by a variable oscillator and a harmonic generator. The complex sounds were constructed on the basis of a $3 \times 3 \times 2 \times 2$ factorial design. The four variables were: (*a*) overall intensity of the stimulus (i.e., the broadband noise with discrete tone components) set at 70 db., 75 db., and 80 db. SPL; (*b*) fundamental tone frequency set at 500 cps, 1,700 cps, and 6,100 cps; (*c*) fundamental tone intensity set at 12 db. and 24 db. SPL above the broadband noise (fundamental tone amplitude relative to $\frac{1}{24}$th octave band); and (*d*) first overtone intensity set at 25% and 75% of the fundamental tone sound pressure level.

The first overtone intensity was determined in terms of a proportion of the fundamental tone intensity, since a relationship between these variables is usually found in noise produced outside the laboratory situation. The 25% and 75% intensities for the 12-db. fundamental tone were 3 db. and 9 db., respectively; and the corresponding intensities for the 24-db. fundamental tone were 12 db. and 18 db., respectively.

Subjects.—Ninety *S*s participated in the experiment to complete a requirement for an introductory course in psychology. Selection was restricted to *S*s who indicated no history of hearing loss and who were naive with respect to the purpose of the study. Two *S*s who failed to follow instructions were replaced with two additional *S*s.

Procedure.—The *S*s were divided into 18 groups of five *S*s each. The conditions were assigned to the groups on the basis of a 3 × 6 factorial design. The two factors were the instructions, varied in three ways, and the order of presentation of the 36 pairs of stimuli, varied in six ways.

The *S*s were brought into the experimental room in groups of five and seated around a table. Low partitions separated each *S*. Each *S* received an 11-point rating scale. The *S*s were given loudness, noisiness, or annoyance instructions. The *S*s in the loudness condition received the following instructions:

> Today I want you to make judgments about the loudness of sounds. When I finish the instructions you are to put on your earphones, I will turn on a tape recorder, and you will hear several pairs of sounds. You are to judge the loudness of the second sound in each pair. Each of you has a rating scale. Notice that each line is numbered from 1 to 11. Under No. 6 is written the word equal, under No. 11 is written much greater, and No. 1 is much less. If the second sound in a pair is equally loud as the first put an X on No. 6. If the second sound is louder than the first put an X on 7, 8, 9, 10, or 11 depending on how much louder it seems to be. If the second sound is less loud put an X on 5, 4, 3, 2, or 1 depending on how much less. It is easiest if you consider each of these numbers as equal units of loudness. The first sound is at Unit 6. Your job is to determine the unit of loudness of the second sound. The sounds will be presented in this way, 3 sec. of Sound 1, 2 sec. of silence, 3 sec. of Sound 2, followed by 10 sec. of silence, during which you are to make your judgments.
>
> Use the first line for the first pair of sounds, the second line for the second pair, and continue on down the page using each line only once. Otherwise, I have no way of determining which pair of sounds your judgments represent. Are there any questions? (pause)
>
> Remember that you are judging the loudness of sounds, the second in each pair, and to use each line only once. If you will put on your earphones I will begin the tape recorder.

The noisiness and annoyance instructions were identical with the exception that words referring to loudness were replaced with the appropriate grammatical form of noisiness or annoyance.

After the instructions were given, *S*s put on earphones, and *E* began a tape recorder. The *S*s rated the second sound in each pair. The first sound in a pair, a broadband noise at 76 db. SPL, was presented to give *S*s a common adaptation level.

When *S*s had completed the ratings they were asked to write their names on the rating sheets. One *S* from each group was selected at random to return for a second session on the following day. During the second session each *S* received the same condition he had received the previous day. The second session was introduced to provide an estimation of the reliability of *S*s' ratings. The experimental sessions lasted approximately 25 min.

RESULTS

The data were taken directly from *S*s' ratings. Pearson product-moment correlations were computed between ratings from the first and second sessions for those *S*s who returned for a second session. The mean correlations, averaged across orders, for loudness, noisiness, and annoyance instructional sets were .90, .85, and .82, respectively. The mean correlation for all groups was .86, which indicated the degree of stability of *S*s' responses. The ratings from the first session were analyzed by a split-plot analysis of variance. Two factors, the six orders of presentation and the three instructional sets, were assigned to *S*s; the levels of the four stimulus factors were assigned within *S*s. The latter variables include: (*a*) three levels of overall intensity, 70 db., 75 db., and 80 db. SPL; (*b*) three fundamental tone frequencies, 500 cps, 1,700 cps, and 6,100 cps; (*c*) two levels of the fundamental tone intensity, 12 db. and 24 db. SPL above the broadband noise level; and (*d*) two levels of the first overtone intensity, 25% and 75% of the fundamental tone SPL.

Order of stimulus presentation.—No differences were found among the six orders of presentation, F (5, 72)

< 1.00, nor was the Order × Instructional Set interaction significant, $F\ (10, 72) = 1.32$, $p > .05$.

The Order × Stimuli interaction yielded a significant F when tested against the Blocks × Stimuli residual error, $F\ (175, 2520) = 2.01$, $p < .01$, (error $MS = 2.09$). This finding indicated that the Order × Stimuli interaction was the appropriate error for tests of significance among the levels of each of the stimulus variables, the interactions among stimuli, and the interactions between set and stimulus variables. Since order was a random variate, each of these effects contained the variance due to the effect itself, the variance due to the interaction between order and the effect, and the error variance. Significance tests of these effects with the residual error term would have been inflated, since the Order × Effect variance had been removed from the residual error. This would have tended to produce a Type I error. In the present study the MS for the Order × Stimuli interaction ($MS = 4.19$, $df = 175$) was used for significance tests of the main stimulus effects and all interactions involving stimulus variables. For significance tests of the two factors assigned between *S*s (i.e., instructional sets, orders, and Order × Instructional Set interaction) the usual within-*S*s error term was used (MS within $S\text{s} = 14.24$, $df = 72$).

Instructional sets.—The means for instructional sets were 5.81 for loudness, 5.86 for noisiness, and 6.83 for annoyance. These differences were significant, $F\ (2, 72) = 5.86$, $p < .01$.

Interactions between instructional set and stimuli.—A clearer picture of the relative emphasis given to each of the sound components is presented by an examination of the interactions between instructional set and stimulus variables. The data pertinent to the Instructional Set × Overall Intensity interaction are presented in Table 1. The ratings under each of the instructional sets were directly related to the overall intensity; however, the variance of the means was greatest under the loudness set, slightly less under the noisiness set, and least under the annoyance set, $F\ (4, 175) = 25.97$, $p < .01$.

TABLE 1

RATINGS OF OVERALL INTENSITY UNDER EACH INSTRUCTIONAL SET

Instructional Set	Overall Intensity Mean Rating			Variance between Means
	70 db.	75 db.	80 db.	
Loudness	3.21	5.69	8.53	7.09
Noisiness	3.26	5.79	8.52	6.92
Annoyance	5.03	7.00	8.46	2.96

The interaction between instructional set and fundamental tone frequency was also significant, $F\ (4, 175) = 7.29$, $p < .01$. The relevant data are presented in Table 2. An examination of the data indicated that the order of the means was the same under loudness and noisiness instructions with the 1,700-cps tone receiving the highest ratings, followed by the 6,100-cps and 500-cps tones, respectively. Relative to the 1,700-cps tone, the

TABLE 2

RATINGS OF FUNDAMENTAL TONE FREQUENCY UNDER EACH INSTRUCTIONAL SET

Instructional Set	Frequency Mean Rating			Variance between Means
	500 cps	1,700 cps	6,100 cps	
Loudness	5.54	6.01	5.87	.058
Noisiness	5.50	6.04	6.03	.095
Annoyance	6.00	7.07	7.42	.547

6,100-cps tone received higher ratings under the noisiness set than under the loudness set. A different order of means was found under the annoyance set. The 6,100-cps tone received the highest ratings, followed by the 1,700-cps tone and the 500-cps tone. These data indicated a shift in emphasis toward the higher frequencies from loudness to noisiness and from noisiness to annoyance instructional sets.

The data relevant to the interaction between instructional set and fundamental tone intensity are presented in Table 3. The interaction was significant, F (2, 175) = 20.56, $p < .01$. Although the 24-db. fundamental tone received higher ratings under each of the instructional sets, the difference between the means for the 12-db. and 24-db. tones was greatest under the annoyance set, followed by noisiness and loudness sets, respectively.

No Instructional Set × Overtone Intensity interaction was found, F (2, 175) < 1.00.

Stimulus variables.—Significant differences were found among the levels of each of the main stimulus variables. Differences among the levels of three of the main effects, however, should be interpreted in terms of the particular instructional set considered. The three variables were overall intensity, F (2, 175) = 5891.70, $p < .01$; fundamental tone intensity, F (1, 175) = 176.87, $p < .01$; and fundamental tone frequency, F (2, 175), $p < .01$.

TABLE 3

Ratings of Fundamental Tone Intensity under Each Instructional Set

Instructional Set	Fundamental Tone Intensity		Diff.
	Mean Rating		
	12 db.	24 db.	
Loudness	5.78	5.84	.06
Noisiness	5.74	5.97	.23
Annoyance	6.27	7.38	1.11

The difference between the levels of the fourth variable, overtone intensity, was also significant, F (1, 175) = 10.02, $p < .01$, with the higher overtone intensity yielding higher ratings than the lower overtone intensity. No significant interaction was found between instructional set and overtone intensity, indicating that the differences were approximately the same under each of the three instructions. The interaction between overall intensity and fundamental tone frequency was significant, F (4, 175) = 2.44, $p < .05$. The differences among fundamental tone frequencies decreased as the overall intensity increased. The data indicated a significant interaction between overall intensity and fundamental tone intensity, F (2, 175) = 3.27, $p < .05$. The differences between the means for the 12-db. and 24-db. fundamental tones decreased as the overall intensity increased. These two interactions were probably due to an increase in masking as the background intensity increased.

The fundamental tone frequency interacted with fundamental tone intensity, F (2, 175) = 8.03, $p < .01$. The data indicated that the difference between the means for fundamental tone intensity was greatest for the 1,700-cps tone, less for the 6,100-cps tone, and least for the 500-cps tone.

The Fundamental Tone Intensity × Overtone Intensity interaction was significant, F (1, 175) = 5.25, $p < .05$. The differences in ratings of the two overtone intensities were greater when presented with the 24-db. fundamental tone than with the 12-db. fundamental tone.

All remaining interactions were not significant.

DISCUSSION

The data indicated that the instructions functioned to set *Ss* to emphasize particular aspects of sound stimuli. Overall intensity was emphasized most under loudness instructions, followed by noisiness and annoyance instructions, respectively. The order of relative emphasis was reversed for fundamental tone frequency and fundamental tone intensity. Highest emphasis was given under annoyance instructions, followed by noisiness and loudness instructions, in that order. In all cases the noisiness instructions yielded data more similar to loudness instructions than to annoyance instructions. No differences in emphasis among instructions were found for overtone intensity.

The differences in emphasis found among instructional sets indicate problems of interpretation of Kryter's (1959, 1960) perceived noise scale. The perceived noise scale could be more accurately referred to as an annoyance scale than as a noisiness scale, since the original research was based on annoyance instructions. The present study indicated that noisiness and annoyance instructions did not yield equivalent results.

The lack of significant second-order interactions among set and stimulus factors suggested that interactions among stimuli were independent of instructional set. If the second-order interactions were in fact zero, the results may indicate that interactions among stimulus variables were due to physiological characteristics of *Ss* since, as such, their form would not be expected to vary under different instructional set.

The results may be applicable to practical problems of community reaction to noise. Until the present time instructional sets have been selected on a more or less intuitive basis. Researchers have assumed that a noise "disturbs" or "annoys" members of a community, and selection of instructions seemed to be based on these assumptions. The findings of the present study indicate an empirical procedure for selection of instructions for scaling human reaction to noise situations. The investigator should first determine the relative contribution of sound components to community reaction. Instructions may then be selected which set *Ss* to emphasize the pertinent sound components. This procedure has the advantage of by-passing the assumption of subjective state inherent in much of the research in this area.

REFERENCES

COPELAND, W. C. T., DAVIDSON, I. M., HARGEST, T. J., & ROBINSON, D. W. A controlled experiment on the subjective effects of jet engine noise. *J. Roy. Aeronaut. Soc.*, 1960,**64**, 33–36.

KRYTER, K. D. Scaling human reactions to the sound from aircraft. *J. Acoust. Soc. Amer.*, 1959, **31**, 1415–1429.

KRYTER, K. D. The meaning and measurement of perceived noise level. *Noise Control*, 1960, **6**, 12–27.

LAIRD, D. A., & COYE, K. Psychological measurements of annoyance as related to pitch and loudness. *J. Acoust. Soc. Amer.*, 1929, **1**, 158–163.

REESE, T. W., & KRYTER, K. D. The relative annoyance produced by various bands of noise. Report No. PB 27306, 1944, United States Department of Commerce, Office of Technical Services, Washington, D. C.

ROBINSON, D. W., & BOWSHER, J. M. A subjective experiment with helicopter noises. *J. Roy. Aeronaut. Soc.*, 1961, **65**, 635–637.

STEVENS, S. S. Calculation of loudness of complex noise. *J. Acoust. Soc. Amer.*, 1956, **28**, 807–832.

(Received April 5, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 221–233

TEMPERATURE ACCLIMATIZATION, RESPONSE STRENGTH, AND THERMAL PREFERENCES IN THE RAT[1]

WARREN H. TEICHNER

University of Massachusetts

3 experiments are reported concerned with the effects of temperature acclimatization on R strength. The 1st dealt with the performance of cold-acclimatized rats in a normal temperature runway when trials having competing Rs are included in or excluded from the average speeds. The 2nd was concerned with the interaction of temperature acclimatization and food deprivation on running speed. Both experiments lend some support to a previous proposal that the metabolic heat production is part of a generalized drive mechanism. The 3rd experiment, performed largely as a control question, scaled the preferences of differently acclimatized rats for air temperature exposures. The results did not detract from the metabolic hypothesis and, in addition, provided specific data on the nature of temperature-preference functions.

Early studies of the effects of temperature acclimatization on runway performance suggested that cold-acclimatized rats run faster in runways of both normal and high temperatures than do normal- and heat-acclimatized animals (Hellmer, 1943; Moore, 1944). Since food was used as the reward for running, and it is known that the food consumption of cold-acclimatized animals is increased, it might be supposed that the increase R strength of the cold-acclimatized *S*s was related to an increased hunger drive. Teichner and Youngling (1962), however, have shown that when essentially satiated animals are used in a shock escape situation, the R strength of cold-acclimatized animals is still greater than that of normals. Thus, another explanation of the effect must be sought.

An interesting possibility, proposed by Teichner and Youngling (1962), relates the known effects of cold acclimatization in animals to the concept of generalized drive (*D*) (Hull, 1943; Spence, 1956). According to this notion, *D* is a hypothetical state of the organism inferred from the relationship between antecedent conditions like food and water deprivation, etc. and the observation that the strength of R varies with the degree of such conditions. *D* is conceived of as a generalized state which (*a*) represents the resultant of all available specific physiological states produced by simultaneously effective antecedent conditions and (*b*) acts as a nonspecific influence on R strength; that is, as *D* varies, R strength varies regardless of the specific nature of the R so that all evoked Rs are affected impartially. Since the metabolic rate is a measure of the general heat-producing activities of all of the cells of the body and since it is known to vary with all specific physiological conditions including those such as food deprivation which have received the most attention in regard to *D*, it is an inviting hypothesis to consider the metabolic rate as the generalized

---

[1] Carried out under Grants NSF G-10918 and NSF G-18632 from the National Science Foundation.

Ann Gustin, Robert LeFrancois, and Leah Price were research assistants.

physiological state for which *D* is a model.

Among the expectations of this approach to *D* are such common findings as those which show that with increased deprivation, the strength of learned Rs increases, general activity increases, etc. Since the effects of cold acclimatization include an increase both in basal metabolic rate and in activity, it is inviting to consider the possibility that one functional effect of such acclimatization is an increase in basal *D* level. Teichner and Youngling's (1962) finding that R strength is increased is consistent with this possibility. The present experiments were an attempt to investigate the utility of a metabolism-*D* concept farther especially in the context of temperature acclimatization.

EXPERIMENT I

Estes (1958) has suggested that, for the rat in a runway, if trials involving Rs in conflict with goal-directed running are eliminated from running speed averages, variations in the time of food deprivation will have no systematic effect on the strength of the running R. Cicala's (1961) results do not support this suggestion, but instead appear clear in showing that with or without including competing R trials in the average running speed, R strength is dependent on the duration of deprivation. If Cicala's results with deprivation could be generalized to the effects of cold acclimatization, some further context of support would be provided for the metabolism-*D* hypothesis. Experiment I was intended to explore this possibility. To do this, conditions permitting still further refinement of the "pure" running R were set up. In addition, the experiment was designed to evaluate the effects of a change in acclimatization temperature on R strength. This was done to extend the results of previous temperature experiments (Hellmer, 1943; Moore, 1944; Teichner & Youngling, 1962) in all of which comparisons were made only as between-temperature effects. Thus, the confounding of temperature and *S* groups remains a possible experimental artifact.

Method

Apparatus.—The apparatus was a straight, wooden runway, 88 in. long, painted flat black inside and out. The top was covered with .5-in. hardware cloth. Manually operated guillotine doors defined the starting and goal boxes each of which was 11 in. long. Three independent, photoelectric sensing, latching relay circuits separated three successive segments of the runway, each 11 in. long, between the starting and goal boxes. The light beam was placed through .5-in. holes whose centers were 2 in. above the floor. Three .01-sec. timers permitted measurement of the time used by *S* to traverse each segment. Visual observations made by *E* furnished records of instances in which competing Rs took place within the limits of each segment. A competing R was defined as any interruption of the forward motion of the animal. It was usually characterized by stopping; retracing; investigating the wall, floor, or upper space; scratching; etc. As long as the forward motion of the animal continued, no competing R was recorded. The *E* recorded only the occurrence of the first competing R within each segment.

Oxygen consumption was measured using a six-channel spirometric apparatus built to the specifications described by Haldi, Wynn, and Breding (1961). Animals were housed in individual cages during Phase I and in temperature-controlled chambers during Phases II and III.

Procedure.—Twenty Sprague-Dawley male rats, 350–400 gm. at the start, were put on a 23-hr. food deprivation regime for 15 days. During this period and the following one they were housed under ordinary conditions in an air-conditioned laboratory maintained at 74° F. In the period which followed (preacclimatization phase), each *S* was run three successive trials per day for 11 days in the runway to a small, constant size, pellet. On the average the animals had been without food for 18 hr. at this time. Oxygen consumptions, body weight, and food consumptions were obtained at periodic intervals up to this point. The first two measures were taken when the animals had been without food for 6–8 hr. Two successive 5-min. O_2 determinations were always taken and averaged to provide a single O_2 estimate.

On completion of the preacclimatization phase, the animals were separated into two groups matched as closely as possible for mean and standard deviation of metabolic rate. One of these groups, chosen at random, was then assigned to continued living at 74° F. The other was assigned to living in 50° F. The two groups were split into two subgroups of five *S*s each and each subgroup was housed in a single, large temperature-controlled chamber. The *S*s were maintained in these cages on the 23-hr. deprivation for 20 days (acclimatization phase) without further exposure to the runway. They were removed only to obtain periodic O_2 consumption measures. Following this (postacclimatization phase) the animals were given three successive daily trials in the runway for 10 days. The runway temperature was 74° F. as before. Otherwise the procedures of the acclimatization phase were continued.

During the preacclimatization phase the mean metabolic rates of the two groups were .54 cc $O_2/cm^2/hr$ for the group continued at 74° F. and .56 cc $O_2/cm^2/hr$ for the group later acclimatized to 50° F. The means of these groups during the postacclimatization period were .46 cc $O_2/cm^2/hr$ for the 74° F. groups and .69 cc $O_2/cm^2/hr$ for the 50° F. group. Thus the preacclimatization matching was quite successful and the two groups differed markedly during the postacclimatization phase. The decrease in O_2 consumption of the 74° F. group is consistent with changes in age and the effects of deprivation to be reported below. Of interest here is only that the increase in the 50° F. group represents an increase above these two factors due to cold acclimatization.

Results

The *E*'s visual observations were used to define trials within segments in which it could be said that at least one competing R had occurred. The mean number of such trials per segment for the pre- and postacclimatization phases is shown in Fig. 1. It may be seen that more competing R trials occurred during the preacclimatization phase, that the number of such trials was not too large, and that the number of such trials decreased from the first through the third segments of the preacclimatization phase whereas later there was little or no difference between the second and third segments. The figure also suggests a slight bias for fewer competing R trials in the 50° F. group in both phases.

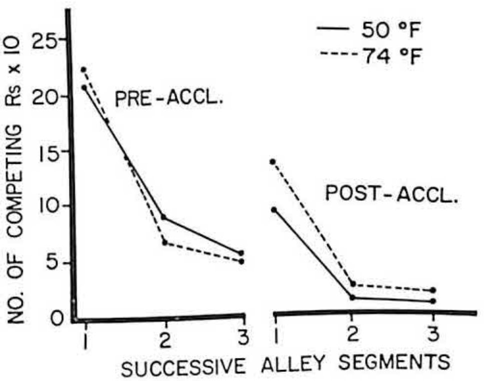


FIG. 1. Number of competing R trials per segment before and after the acclimatization period.

Figure 2 presents the mean reciprocal running times taken over each phase for the two groups in the three segments of the runway. This figure also permits a comparison of the data when all trials are averaged in and when trials having competing Rs are omitted from the average. Again the data show a slight, but consistent bias of the 50° F. group for greater speeds. They suggest too that when competing Rs are omitted during the preacclimatization phase, the bias is increased. In spite of this, these results appear clear in suggesting that the running speed of the cold-acclimatized group tended to increase more than that of the 74° F. group.

Inspection of Fig. 2 suggests that running speed tended to be greatest in the middle segment. This is not consistent with the goal-gradient

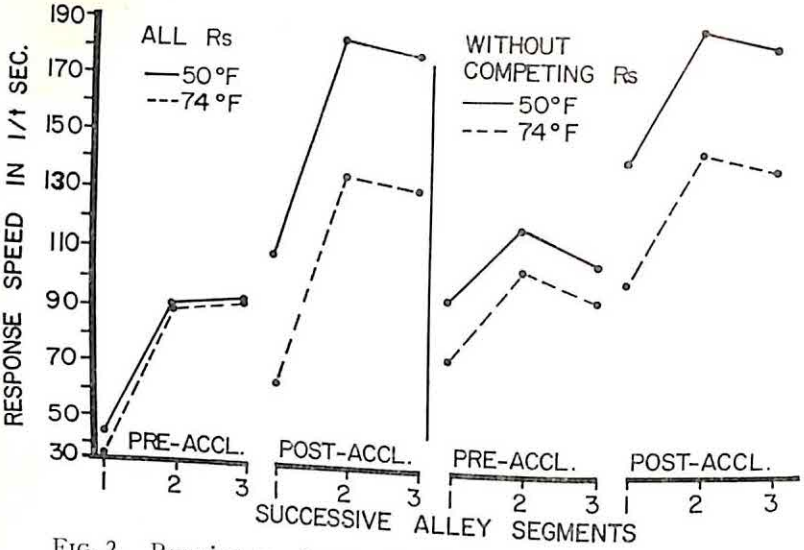


FIG. 2. Running speed averages per segment with and without competing Rs, before and after the acclimatization period.

principle (Hull, 1943), but is reasonable in light of the requirement for the animal to decelerate as it approached the goal. In the same way, the failure of the first segment to yield speeds as high as the second was presumably due to the initial acceleration required. When these things were considered along with the very small differences in the number of competing Rs shown in the last two segments in Fig. 1, it appeared reasonable to select the middle segment as providing the best measure of a "pure" running R. For this reason the second segment was selected for more detailed analysis.

Figure 3 presents the mean reciprocal running time per day in the second segment for the conditions of the experiment. It is apparent from these data that no difference in interpretation of the results is permitted by averaging all trials as compared to omitting competing R trials. The figure also suggests that the small bias in running speed shown in Fig. 2 was not consistent over the preacclimatization phase. In fact, considering the all R data, the 50° F. group was slower than the 74° F. group in the last 5 days of this phase and, omitting competing R trials, the differences between the groups were very small and inconsistent during the same period. On the other hand, considering the postacclimatization period, the data are clear in revealing a consistently greater running speed for the 50° F. group at all times.

Since there were no essential differences in the speed data with or without including competing R trials,

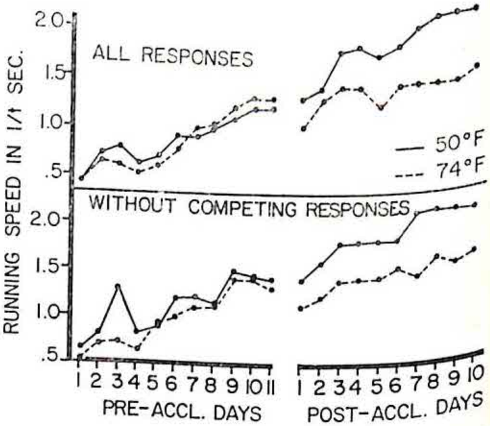


FIG. 3. Mean daily running speeds for the center segment with and without using competing R trials, before and after the acclimatization period.

analysis of variance of the reciprocals was performed only on the all R data. The results of the analysis of the preacclimatization data are shown in Table 1. As may be seen the analysis confirms Fig. 1 in suggesting no differences between the two groups during this phase. The analysis did not indicate that there was a difference in running speed on the three trials within each day. It did support the rising acquisition trends of Fig. 3 by a significant Days effect and it supported the indication in Fig. 2 that running speeds differed according to the segment of the runway measured. The lack of a significant Groups × Segments interaction supports the interpretation that the differences between groups shown in Fig. 2 during this phase did not constitute inadequate matching.

TABLE 1

Analysis of Variance of Preacclimatization Reciprocal Times from the Center Segment Using All Trials

Source	*df*	*MS*	*F*
Between total	17		
Climate (C)	1	.86	.13
Ss/C	16	6.72	
Within total	1,620		
Days (D)	9	8.60	20.25***
Segments (P)	2	43.66	121.00***
Trials (T)	2	1.32	3.17
C×D	9	.28	.66
C×P	2	.13	.35
C×T	2	.25	.60
D×P	18	.76	8.10***
D×T	18	.19	1.18
P×T	4	.52	6.51***
C×D×P	18	.04	.47
C×D×T	18	.20	1.23
D×P×T	36	.08	1.02
P×T×C	4	.04	.50
C×D×P×T	36	.10	1.27
Ss×D/C	144	.42	
Ss×P/C	32	.36	
Ss×T/C	32	.42	
Ss×D×P/C	288	.09	
Ss×D×T/C	288	.16	
S×P×T/C	64	.08	
Ss×D×P×T/C	576	.08	
Total	1,620		

*** $p < .001$.

TABLE 2

Analysis of Variance of Postacclimatization Reciprocal Times from the Center Segment Using All Trials

Source	*df*	*MS*	*F*
Between total	17		
Climate (C)	1	58.37	5.70*
Ss/C	16	10.24	
Within total	1,620		
Days (D)	9	8.91	21.23***
Segments (P)	2	87.40	293.39***
Trials (T)	2	.28	.75
C×D	9	.38	.91
C×P	2	.01	.04
C×T	2	2.42	6.45**
D×P	18	.20	2.23***
D×T	18	.12	.62
P×T	4	.38	5.32***
C×D×P	18	.11	1.21
C×D×T	18	.14	.75
D×P×T	36	.08	1.12
P×T×C	4	.02	.31
C×D×P×T	36	.04	.50
Ss×D/C	144	.42	
Ss×P/C	32	.30	
Ss×T/C	32	.37	
Ss×D×P/C	288	.09	
Ss×D×T/C	288	.19	
S×P×T/C	64	.07	
S×D×P×T/C	576	.07	
Total	1,620		

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 2 summarizes the analysis of variance performed on the all R postacclimatization data. In support of the graphical findings, the table shows a significant group difference. Otherwise the main effects were the same as during the preacclimatization phase and the interactions were not importantly different. Taking all the results together, Exp. I appears to permit the conclusion that an increase in R strength was associated with cold acclimatization and that this was true regardless of whether trials containing competing Rs were included in the data or not.

EXPERIMENT II

As animals are deprived of food and their food stores depleted, the resting metabolic rate will decrease. If it is assumed that the metabolic rate has an effect on R strength, then the strength of a learned R should decrease accordingly. On the other hand, it may be possible to compensate for the decrease in metabolic rate resulting from deprivation by acclimatizing animals to different temperatures. Thus, the possibility exists of investigating the joint effect of these physiological conditions, temperature and deprivation, on metabolic rate and then of determining whether the effects of the metabolic rate on R strength depend only upon the resultant metabolic value or whether they are specific to the separate conditions on which the metabolic rate depends. This experiment was a first attempt along this line.

Method

The experimental design was a 2 × 4 factorial of temperatures and deprivations. Eighty male Sprague-Dawley rats, 93 days old and approximately 350 gm. weight at the start, were randomly assigned to eight groups of 10 *S*s each made up of living temperatures of 45°, 65°, 80°, and 90° F. in combination with a food deprivation regime of 6 or 18 hr. The *S*s lived in groups of 5 under their assigned conditions for 15 days prior to experience in the apparatus.

The deprivation regimes were accomplished by keeping *S*s on ad-lib. feeding until either 6 or 18 hr. before the time scheduled for use in the apparatus. Following this time (about 1 min.), *S*s were fed individually in a device which partitioned them within the temperature chamber. They were permitted 90 min. of individual feeding following which they were rejoined as a group and permitted ad-lib. feeding until the appropriate time for removing the food. Individual food consumptions and body weights were obtained daily throughout the experiment.

The apparatus was a 33-in. runway all but the center 11 in. of which was made up of a starting and goal chamber. The runway temperature was 74° F. Measures of the time required to traverse the center 11-in. section were obtained with a photoelectric circuit and an electronic timer. On each trial *S* found a large, hard food pellet in the goal. He was permitted 45 sec. of eating time in the goal box and the amount of the reward consumed was determined. All food weighings were made with a triple-beam balance accurate to .01 gm. Following the 15-day acclimatization period, *S* was given one trial per day in the runway for 8 successive days. All prerunway living and feeding conditions were maintained.

Oxygen consumptions were obtained by the procedures described above. These measures were taken on the last 11 days of the prerunway period and on the day following the runway period always at the same time of day at which the runway experience was given.

Results

Analyses of variance were performed on the body weight, food consumption, reward consumption, and the reciprocal running time for the 8 days in which the runway was used. These analyses are summarized in Table 3. To analyse the O_2 consumptions converted to metabolic rates, the 12 days on which these were obtained were used. This was necessary since no data were available for the period in the runway. The results of this analysis are shown in Table 4.

Inspection of these two tables shows that the food consumptions were stable over the runway period. On the other hand, all other measures varied significantly throughout this period. Plots of these results indicated that running speed and body weight increased while food consumption and metabolic rate decreased.

Tables 3 and 4 also show that all measures were significantly influenced by temperature and that all but running speed were significantly influenced by the different deprivation times. Running speed, amount of reward consumed, and metabolic rate

TABLE 3

Analyses of Variance of Body Weight, Food Consumption, Reward Consumption, and Reciprocal Running Time for Exp. II

Source	*df*	Body Weight		Food Consumption		Reward Consumption		Speed (1/t)	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between *S*s									
Temperature (T)	3	69,887.0	19.9**	156.0	7.83**	.161	9.2**	.41	2.74*
Deprivation (d)	1	197,297.0	56.2**	5,085.0	255.24**	.294	16.8**	.51	3.41
T × d	3	2,837.0	.8	41.0	2.07	.162	9.3**	.72	4.82**
*S*s/T × d	72	3,512.0		20.0		.017		.15	
Within *S*s									
Days (D)	7	3,123.0	64.3**	7.7	1.96	.018	6.7**	.27	4.16**
T × D	21	92.0	1.9**	7.2	1.85*	.011	4.0**	.06	.92
d × D	7	111.0	2.3*	15.2	3.90**	.018	6.5**	.04	.59
T × d × D	21	115.0	2.4**	7.9	2.02**	.011	4.0**	.08	1.23
S × D/T × d	504	48.0		3.9		.003		.06	

* $p < .05$.
** $p < .01$.

were related to a significant temperature-deprivation interaction whereas this was not the case for body weight and food consumption.

Table 5 presents the main effects of temperature and deprivation on body weight, food consumption, and reward consumption in terms of the mean per group for the last 5 days of the runway period. The table shows that within each deprivation group, body weight tended to increase with increasing temperature until 90° F. at which there was a reversal. The body weights of the 18-hr. group were uniformly the smaller. The food consumption data do not follow a similar trend although there is some suggestion of a decreasing consumption with increasing temperature. Both deprivation groups were consistent in showing inversions in such a trend at the same temperatures.

The reward consumption data indicate that very little of the reward was consumed by either group although the amount consumed by the 18-hr. group was consistently greater. Actually, since the mean values shown are smaller than the accuracy of the measuring instrument which was used, the data do not suggest that any important quantity of reward was consumed by any of the 6-hr. groups or by the 18-hr. group acclimatized to 80° F. In fact, the values shown which are less than .01 gm. in all cases were due to one *S* within each group and then on only 1 day of the 5.

Figure 4 presents the effects of temperature and deprivation on run-

TABLE 4

Analysis of Variance of the Metabolic Rates (Cc $O_2/Cm^2/Hr$) of Exp. II

Source	*df*	*MS*	*F*
Between *S*s			
Temperature (T)	3	1.73	3.14*
Deprivation (d)	1	8.83	16.00**
T × d	3	3.36	6.09**
*S*s/T × d	72	.55	
Within *S*s			
Days (D)	11	.59	1.90*
T × D	33	.22	.71
d × D	11	.94	3.05**
T × d × D	33	.17	.55
S × D/T × d	792	.31	

* $p < .05$.
** $p < .01$.

TABLE 5

Mean Body Weight, Food Consumption, and Reward Consumption for the Last 5 Days of Runway Training in Exp. II

Deprivation (Hr.)	Temperature (Degrees F.)	Body Weight (Gm.)	Food Consumption (Gm.)	Reward Consumption (Gm.)
6	45	342	5.4	.001
	65	363	3.8	.003
	80	384	4.1	.000
	90	355	3.7	.000
18	45	291	11.2	.198
	65	332	9.4	.019
	80	354	10.2	.004
	90	322	7.7	.256

ning speed and metabolic rate. The former, as before, is the mean of the last 5 days. Since the metabolic rates apparently were not affected during the runway period, the mean of the last 5 days on which these measures were obtained was plotted. Four of these days preceded the runway period and the fifth followed it by 1 day.

This figure shows that the metabolic rates of the 18-hr. groups were always lower than their corresponding 6-hr. groups. In general, the metabolic rates were inversely related to the temperature of acclimatization although the 18-hr. group showed a reversal at 90° F. For the 6-hr. groups, running speed was also inversely related to temperature. For the 18-hr. groups the running speed trend was reversed after 65° F. so that for these groups running speed was first inversely, then positively related to acclimatization.

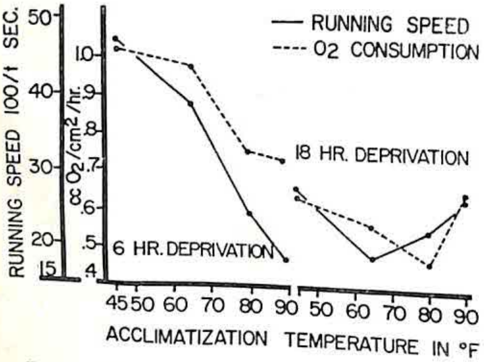


Fig. 4. Running speed and metabolic rate as a function of acclimatization and deprivation.

Discussion

Comparing the two deprivation groups in regard to performance, it may be seen that the 6-hr. groups ran faster in all cases except for those animals acclimatized to 90° F. This general result, i.e., better performance with the lesser deprivation, has some theoretical basis of expectation in terms of the effects of the multiplicative relationship between habit strength and generalized drive postulated by Hull (1943). As noted by Brown (1961, p. 106), according to this postulate, performance will be impaired at high levels of drive whenever the strongest R association is an incorrect one. It seems reasonable to assume that early in learning, Rs other than running to the goal are dominant in the runway situation. Accordingly, the 18-hr. group would be expected to show slower running speeds early in learning, although faster speeds later. Since the experiment dealt only with early learning, only the first part of this expectation can be evaluated. Figure 4 offers some support of the expectation.

Although the results support the expected better performance of the 6-hr. group early in learning, the lack of a significant main deprivation effect does not. Neither does it support an opposite expectation. This result, lack of a significant deprivation effect early in learning, is consistent, however, with other recent deprivation studies (Barry, 1958; Besch & Reynolds, 1958; Brown & Belloni, 1963). The effect is not without precedent, therefore, and lends credence to the significant Temperature × Deprivation interaction which was obtained. Inspection of the figure suggests, at least for early learning, that the effects of temperature acclimatization depend upon the duration of food deprivation such that for short deprivations, a monotonic, inverse temperature function obtains within the range studied whereas for longer deprivations, the function may be U shaped. This result is important to the basic interest of the experiment since the metabolic rates, crudely, showed similar relationships. Other things equal, the data appear to suggest some support for an interpretation of metabolic rate as a primary R-strength determining factor. Furthermore, within the same context, the failure of the 18-hr. group to show greater R strength suggests that, at least early in learning, the effect of metabolic rate may not be dependent upon the specific conditions on which it depends. These possibilities, however, need further study.

EXPERIMENT III

In setting up the previous experiment, it had been hoped that the amount of food reward consumed might be relatable to other measures of the experiment. The failure of so many of the groups to accept the reward was unexpected especially since all groups showed a systematic increase in running speed over the 8-day period. Presumably, secondary reinforcements were involved in establishing the acquisition curves, e.g., removing and handling *S* following each run, but it is not clear how such reinforcements might have differed in magnitude systematically enough to produce the systematic relationships among the 6-hr. groups none of which could be said to have eaten the reward.

A possible bias in the second experiment might have resulted from the fact that the runway temperature was always 74° F. Thus, the difference between this temperature and that to which *S*s were acclimatized might have been related to a thermal preference as a basis for differential reinforcement. Experiment III was designed, therefore, to study the effect of acclimatization on thermal preferences during exposures approaching the duration of exposure used in a variety of previous experiments as well as the last one. In addition, the nature of the experiment was such as to provide data relevant to the effects of age on the kinds of physiological measures taken in Exp. II.

Method

Thirty Sprague-Dawley rats, 80–85 days old, 275–305 gm. body weight, were separated into six groups of 5 of approximately equal mean and range of body weight. Each group was housed in a thermally controlled cage maintained at 45°, 55°, 80°, 85°, 90°, or 95° F., 45% relative humidity. Prior to this all animals had been living in 74° F., 45% relative humidity. Ad-lib. feeding and drinking were permitted at all times. For the first 47 days the animals were permitted to acclimatize. Daily measurements were made during this period of body weight, food consumption, and O_2 consumption. These measurements were repeated on Days 70, 98, 118, and 138.

During Days 48–138 the animals were tested for ambient temperature preferences. The apparatus was a 6.0 × 4.5 × 19.0 in. (inside dimensions) rectangular box of insulated, 1-in. double-walled plywood construction. A 1-in. thick center partition containing a vertically swinging door divided the box into two compartments, each 9 in. in length. The door permitted the animal easy passage from one compartment to the other. Each compartment was maintained at a different mean air temperature. The heat sources were shielded electric elements located at the far end of each compartment (end away from center). Slowly moving air was passed from blowers across the elements and then baffled prior to entering the compartment. The roof of the box could be removed and thus served to insert or remove the animal from the apparatus. The roof was constructed of aluminum and on its outside surface held two insulated, wooden containers built as superstructures into which were placed dry ice as necessary to cool the compartments. The heating system worked against this cooling to maintain any desired internal temperature setting. A thermistor-sensing, electronic thermoregulating system maintained mean compartment temperature within 1° F. at all settings.

After the animal was placed into one of the compartments, a period of 30 sec. was permitted to elapse before beginning a 6-min. measuring period. A photoelectric relay circuit operated by movements of the swinging door controlled .01-min. timers which accumulated the time spent by each animal in each compartment. The compartment in which more than half of the 6-min. period was spent was taken as the preferred one. Each animal made two "judgments" per day approximately 5 hr. apart.

By maintaining the compartment temperatures differentially, a complete paired-comparison procedure was carried out for every combination of 68°, 72°, 76°, 80°, 84°, 88°, 92°, 96°, and 100° F. The order of the comparison was random and the same for all groups. The entire paired-comparison series was replicated four times using the same random sequence for the first and third replications and the reverse of this sequence for the second and fourth. Across replications, the temperature of each pair into which the animal was first placed was counterbalanced as was the compartment into which it was first placed.

Results

An analysis of variance was performed on each of the physiological measures. All of them, that is, body weight, food consumption, and O_2 consumption, varied significantly with acclimatization conditions and with experimental days ($p < .001$). The interaction between these two variables was also significant ($p < .001$) in each case. Since the total experimental time of each animal was approximately 5 mo., the variable, experimental days, represents the effects of age from approximately 3 to 8 mo.

To evaluate the temperature preferences, a paired-comparison scale was constructed for each of the four replications for each of the acclimatization groups. The first replication exhibited considerably more variability than the others. It was discarded, therefore, as being an instructional period for the animals. Analysis of variance of the remaining scale values indicated that both comparison ($p < .001$) and acclimatization ($p < .05$) temperatures as well as their interaction ($p < .001$) were significant variables.

Plots of the various dependent measures as a function of acclimatization temperatures revealed a considerable amount of overlap between successive acclimatization groups. Since *S*s lived in groups within the temperature chambers, it was possible that the small differences in acclimatization temperatures used (as small as 5° F. ± 2° F.) were contaminated to the point of overlap. Attempts to study this with the use of the Duncan range test were helpful only in suggesting that the safest comparisons were most likely to be among the 45°, 80°, and 95° F. groups. For this reason the data obtained from the remaining intermediate groups were not used for further analysis.

A summary of the mean changes per group in food consumption, body weight, and metabolic rate over the course of the experiment is presented in Table 6. Food consumption relative to body weight can be seen to have decreased consistently over the 5-mo. period even though body weight increased. A plot of the body weight data in Table 6 reveals that the 95° F. group increased its weight linearly and

TABLE 6

Mean Food Consumption, Body Weight, and Metabolic Rate for Different Acclimatization Temperature Groups during Exp. III

Days	Relative Food Consumption (Gm Food/Gm Body Weight)			Body Weight (Gm.)			Metabolic Rate (Cc O_2/Cm^2/Hr)		
	45	80	95	45	80	95	45	80	95
1–11	.13	.09	.08	300	308	275	.81	.67	.62
11–22	.13	.07	.08	340	338	291	.74	.55	.53
23–33	.11	.06	.07	371	366	362	.75	.62	.52
34–44	.09	.06	.06	399	405	333	.75	.61	.31
98–138	.08	.05	.04	492	481	421	.59	.45	.35

Note.—The age of the animals on Day 1 was 80–85 days.

that the two lower temperature groups increased in weight in an essentially linear fashion through the first four points, but that the increase to the last point decreased slightly. The same plot also shows that the body weight increases of the 45° F. and 80° F. groups were the same and that both were greater than the high-temperature group. On the other hand, the food consumptions of the two higher temperature groups were the same and both were consistently below that of the cold group. Thus, the data suggest that the effect of cold acclimatization relative to heat acclimatization is an increase in both body weight and food consumption, but that relative to 80° F. the effects of the cold are better distinguished by food consumption whereas the effects of the heat are better differentiated by body weight. Both measures change systematically with age, with food consumption being inversely related and body weight directly related.

Inspection of Table 6 shows that metabolic rate also changed with age. A plot of these values indicates that metabolic rate decreased for all groups from the first to the second period, then leveled off until the third period and from that point declined again to the last point. Unlike either food consumption or body weight, this measure provided a clear differentiation among all three acclimatization groups. Over the entire experiment, the metabolic rates of the groups were higher the lower the temperature.

Figure 5 presents the thermal preferences expressed as the mean of the last three replicates. To provide a basis for comparing scales, the obtained scale values were converted to their percentage positions on the scale ranges (Woodworth & Schlosberg, 1954). Inspection of the figure shows that the 45° F. group preferred higher temperatures less than did the 80° F. group. The preferences of these two groups were quite different from those

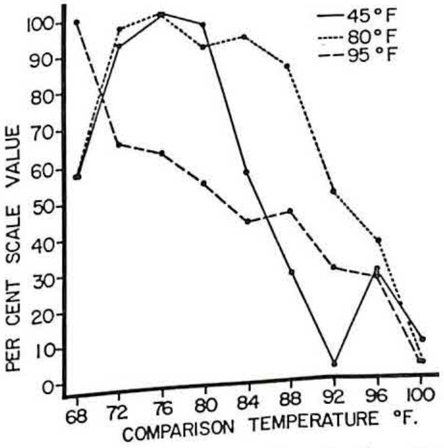


Fig. 5. Effects of acclimatization temperature on preference for short exposures of air temperature.

of the 95° F. group. This latter group exhibited no meaningful inversion over the comparison range, i.e., it showed no optimum temperature. Instead, the preferences of the 95° F. group were greater the lower the temperature. Presumably, at some low temperature the preferences of this group would decrease, but within the range studied this did not happen.

Discussion

The results of the experiment do not support the hypothesis that thermal preferences, within the normal living range, are affected by cold acclimatization. They do suggest that heat acclimatization modifies these preferences within this range, but the effect is not an increased preference for warmer temperatures, but just the opposite, i.e., increased preferences for lower temperatures. These results, of course, apply only to short exposures.

General Discussion

The results of the first experiment appear clear in showing that the effect of cold acclimatization is an increase in R strength and that this is true regardless of the inclusion of competing R trials in the average speeds. Results of this kind, using deprivation as the independent variable, have been used as support for a *D* concept (Cicala, 1961; Spence, 1956). As a contextual inference, temperature acclimatization may be also supposed to fall into the class of independent variables on which *D* depends.

The results obtained with the 6-hr. deprivation group of Exp. II also support the expectation of an inverse relationship between the temperature of acclimatization and the strength of R. Experiment III provides no support for an explanation of this relationship in terms of differential air-temperature preferences. Since, as noted, there is no basis for assuming a differential secondary reinforcement among the various groups, these results may also be considered in terms of variations in *D* associated with the conditions of acclimatization. Since these results and those of Exp. I demonstrated that both R strength and metabolic rate tend to be inversely related to acclimatization temperature, some further credence may be given to the proposal of Teichner and Youngling (1962) that the metabolic-heat production plays an important role as part of the physiological mechanism represented by Hull's (1943) hypothetical intervening construct, *D*. More direct, less correlationally inferred evidence should be sought.

The results of Exp. III have additional interests. First, they do not suggest that the temperature preference of cold-acclimatized animals is altered whereas those of heat-acclimatized animals do change. These effects did not appear related to the metabolic heat production. When this is considered along with the finding that operant Rs can be established for cold-exposed rats using "shots" of heat as a reinforcement (e.g., Carlton & Marks, 1958; Weiss & Laties, 1961), it would appear that the preferences are likely to be related to the rate of surface heat loss. The same consideration might be important to the problem of R strength as studied in the first two experiments. More specifically, the heat balance (rate of heat production/rate of heat loss) may be more descriptive of the generalized physiological state. Further studies should consider the measurement of additional parameters of body-heat regulation.

A second aspect of the results of Exp. III of interest concerns the effects of age and temperature on measures which are conventionally used as a basis for inferring strength of hunger (or thirst) drive. The results are clear in showing again that both body weight and food consumption depend on these factors. However, it was also found that with increasing age, these two measures are inversely related. That is, with age rats get heavier and eat less. Thus the choice of either one as a basis for inferring drive strength in deprivation studies must necessarily involve a confounding of age with the conditions of deprivation. *On* the other hand, although the metabolic rate also decreases with age, it depends upon food consumption and, using O_2 consumption, it is calculated from a consideration of body weight. Thus, all of these variables are represented by a single resultant value. In addition, the resultant metabolic rate includes the effects of temperature which are not represented as consistently by the other two measures.

REFERENCES

Barry, H., III. Effects of strength of drive on learning and extinction. *J. exp. Psychol.*, 1958, **55**, 473–481.

Besch, N. F., & Reynolds, W. F. Alley length and time of food deprivation in instrumental reward learning. *J. exp. Psychol.*, 1958, **56**, 448–452.

Brown, J. S. *The motivation of behavior.* New York: McGraw-Hill, 1961.

Brown, J. S., & Belloni, M. Performance as a function of deprivation time following periodic feeding in an isolated environment. *J. comp. physiol. Psychol.*, 1963, **56**, 105–110.

Carlton, P. L., & Marks, R. A. Cold exposure and heat reinforced behavior. *Science*, 1958, **128**, 1344.

Cicala, J. A. Running speed in rats as a function of drive level and presence or absence of competing response trials. *J. exp. Psychol.*, 1961, **62**, 329–334.

Estes, W. K. Stimulus response theory of drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: Univer. Nebraska Press, 1958.

Haldi, J., Wynn, W., & Breding, H. Apparatus for measuring oxygen consumption of small animals. *J. appl. Physiol.*, 1961, **16**, 923–925.

Hellmer, L. A. The effect of temperature on the behavior of the white rat. *Amer. J. Psychol.*, 1943, **56**, 408–421.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Moore, K. The effect of controlled temperature changes on the behavior of the white rat. *J. exp. Psychol.*, 1944, **34**, 70–79.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Teichner, W. H., & Youngling, E. Acclimatization, habituation, motivation, and cold exposure. *J. comp. physiol. Psychol.*, 1962, **55**, 322–326.

Weiss, B., & Laties, V. G. Behavioral thermoregulation. *Science*, 1961, **133**, 1338–1344.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, 1954.

(Received April 10, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 234–241

POSITIVE AND NEGATIVE GRADIENTS OF RESPONSE STRENGTH IN A TEMPORAL CONFLICT SITUATION[1]

JOHN LEE WIPF[2]

University of Florida

The assumption was tested that strengths of tendencies to respond to positive and negative aspects of a future ambivalent event vary in time in the manner that approach and avoidance tendencies change in a spatial conflict situation. On each of 116 trials, human Ss made a disjunctive manual response to 1 of 2 stimuli. A stimulus was presented every 12 sec. except on 16 aperiodic test trials, when the signal occurred with reduced intensity at 3, 6, 9, or 12 sec. This procedure was used to show temporal gradients of response strength to positive and negative stimuli (30 Ss), to positive and neutral stimuli (30 Ss), and to negative and neutral stimuli (30 Ss). Gradients were established; the negative (avoidance) was steeper and higher than the positive (approach).

Brown (1957) described a temporal approach-avoidance conflict situation as one in which the goal is a future event, with both positive and negative aspects, to which *S* has competing response tendencies. In the context of Miller's (1944, 1959) formulations, it would be predicted that the variations in the strength of tendencies to respond to the generalized rewarding and punishing aspects of the future event, with nearness in time to that event, would resemble the changes in approach and avoidance tendencies in a spatial conflict situation.

Several investigators have reported that a response repeatedly elicited at a regular time interval may become conditioned to the cues characteristic of that interval, and that the strength of the response when tested at various shorter time periods increases as the customary time of stimulus presentation is approached. The gradient of increasing response strength established in this way is usually termed a *temporal gradient.* Brown (1939) shocked rats at regular 12-sec. intervals and then demonstrated a temporal gradient of strength of unconditioned jumping responses by introducing shocks at shorter (and longer) time intervals; the strength of the response decreased as the test intervals became shorter. Mowrer (1940) used a similar temporal paradigm with human *S*s, and established a reaction-time gradient to tones presented at time intervals shorter (and longer) than the conditioned 12-sec. interval. Again, response strength decreased as a function of decreasing test-interval length. Although Brown's results may be thought to show a temporal gradient of avoidance, and Mowrer's as showing an approach-like temporal gradient, there has been only one study specifically aimed at corroboration of conflict principles in a temporal situation (Rigby, 1954). Using a somewhat different temporal conditioning procedure, Rigby demonstrated temporal gradients of approach and avoidance, but failed to confirm the hypothesis of a steeper avoidance gradient.

---

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of Florida. The author is affectionately grateful to Judson S. Brown for his encouragement and critical counsel as dissertation advisor. This research was supported in part by Grants NIH-M-2721-C2 and NIH-M-4952.

[2] Now at University of Cincinnati.

The present experiment was designed to demonstrate temporal gradients of approach and avoidance of differential steepness. Strictly speaking, in the temporal situation used, "approach" and "avoidance" might more appropriately be referred to as tendencies to respond to positive and negative aspects, respectively, of an anticipated event. The gradients of generalized response tendencies, therefore, are henceforth called the positive and negative gradients.

The general procedure bore a close resemblance to the temporal paradigm employed by Mowrer (1940). On each trial, human *Ss* were required to make a simple manual response to one of two stimuli presented singly in balanced random order at regularly recurring 12-sec. intervals. Aperiodically in a sequence of 12-sec. trials, the stimuli were introduced with reduced intensity after intervals of 3, 6, 9, or 12 sec. In one experimental condition electric shocks and tones were used as the cues to respond. The tones were made positive by monetary reward, and monetary loss was used to increase the negative value of the shock. This conflict condition was designed to permit the simultaneous establishment of positive and negative temporal gradients. To a second experimental group positive tone signals were presented in the same random sequences as in the conflict group but with neutral light stimuli replacing the shock response cues. The positive gradient established in this group could then be compared to the positive gradient from the conflict condition. A third group was used to establish a negative gradient comparable to the negative gradient of the conflict group. Again the loss of money gave additional negative value to the shock stimuli and neutral light signals were used to counterbalance the sequence.

METHOD

Subjects.—The *Ss* were 60 male and 30 female undergraduate psychology students at the University of Florida. They were assigned randomly to three equal-sized groups. The proportions of males and females and of right- and left-handed *Ss* were held constant between groups.

Apparatus and procedure.—The *S* was seated at a standard-height table in a black-curtained cubicle illuminated by a shaded 7.5-w. red lamp. About 5 min. were allowed for *S* to adapt to the reduced level of illumination while being familiarized with the response panel, instructions, and signals.

The response panel consisted of a black plywood board (32 × 24 in.) elevated at the back (6 in.) and sloping toward *S*. Protruding from the surface of the board were three black buttons connected by short shafts to microswitches mounted under the board. These response keys were placed 9 in. apart in the form of an inverted triangle, with the key at the apex 8 in. from the front edge of the board. Centered behind each key was a white "business" card providing a label. For all groups the front key was labeled "rest." The two back response keys were labeled "+" (tone), "−" (shock), and "0" (light), in the combination appropriate to *S*'s experimental group.

The data-collection period consisted of 116 trials. On each trial one of two signals was presented, tone or shock for the conflict group (Group C), tone or light for the positive group (Group P), and shock or light for the negative group (Group N). When the signal came on, *S* was required to respond as quickly as possible by lifting the finger tips of his preferred hand from the rest key and pressing the response key paired with that signal. Key-signal pairings were counterbalanced between *Ss*, sex, and handedness, for control purposes. For any one *S*, response to a particular signal was always to the same key. Pressing the appropriate response key terminated the signal.

Group C *Ss* were given $1.00 initially and were told that each shock signal indicated the loss of up to 1 nickel and each tone signal meant the gain of up to 1 nickel. They were instructed that the faster the response to the negative signal, the smaller the amount they would lose and the faster the response to the positive signal, the greater the amount they would win. The *S*s in Group P were given no money to start with, but were assured that with each tone signal they could win up to 1 nickel depending on their speed. The light signal did not indicate the gain or loss of anything. Group N *S*s started with $4.00 and were told that they would lose up to 1 nickel for each shock but that the faster the response the smaller the loss. The light signal was indicated as neutral as in Group P. To maximize cooperation, *S*s were shown money receipts signed by previous *S*s.

For Groups C and P, the termination of the tone was followed immediately by the sound of money dropping into a coin box. This sound was produced when a solenoid, momentarily activated by *S*'s turning off the tone, jerked a small tin can containing some coins. The can was suspended from the solenoid which was located next to the experimental cubicle.

Signals to respond were presented every 12 sec. (i.e., 12 sec. after the onset of the signal of the preceding trial), except on Trials 11, 16, 24, 31, 39, 48, 54, 60, 67, 73, 81, 90, 95, 102, 110, and 116. On these *test* trials, the signal came on after intervals of 3, 6, 9, or 12 sec. Test Trials 11, 48, 67, and 95 occurred 3 sec. after the beginning of the previous trial; Trials 31, 54, 73, and 116 after 6 sec.; Trials 16, 24, 81, and 110 after 9 sec.; and Trials 39, 60, 90, and 102 after 12 sec. The intensity of the test-trial signals was reduced to minimize the dynamogenic contribution of the stimulus to the response speed and to maximize the effect of generalized stimulus and temporal cues.

Each of the two signals was presented twice at each of the four test intervals, totaling 16 test trials. The data reported were collected on these trials. Within the first and second eight test trials, each signal occurred at each interval. This dichotomy, although arbitrary, permitted statistical comparison of the first half of the data with the second half, but there was actually no break in the continuity of the procedure. The trial immediately following a test trial occurred at the standard 12-sec. interval after the onset of the test trial. Immediately prior to any test trial, the frequencies of the previous occurrences of the two signals were equal or differed only by one.

Ten training trials preceded the first test trial, and thereafter an average of 6 training trials preceded each test trial. Two sequential orders of presentation of signals were used. Half of each group received the stimuli in a random sequence developed with the restrictions made necessary by the various control and balancing procedures mentioned above. The other half received the signals in the converse order; for example, on the first 10 trials, half of Group C received shocks and tones in the order STSSTTSSTT, and the other half received TSTTSSTTSS.

The shock signals were delivered through $\frac{3}{4}$-in. silver electrodes to the balls of the index and middle fingers of *S*'s nonpreferred hand. The shock voltage was produced by passing a 117-v., 60-cycle alternating current through a variac and two isolation transformers. The shock intensities employed were 42.5 v. on the training trials and 25 v. on the test trials. These voltages were measured across the electrodes with a vacuum-tube voltmeter without *S* in the circuit.

The tone signals were 1,000-cps tones generated by a General Radio beat-frequency oscillator, Type No. 1304-A. The output control was set to produce a 1-v. signal. This signal was passed through a General Radio decade attenuator which delivered it to a set of Trimm, ANB H-1, headphones with MX-41/AR cushions. The headphones had been calibrated previously to determine that each phone delivered approximately 100 db. when 1 v. was measured across it (all decibels re: .0002 dynes/cm^2). The two tone intensities used were obtained by attenuating the signal to 76 db. (strong tone) and 66 db. (weak tone).

The light-signal source was a 50-w., 117-v. frosted incandescent bulb located above and behind *S*. The weakened signal used on test trials was obtained by introducing a hidden 50-w., 117-v. lamp in series with the signal light, thus reducing its intensity. These signals provided an overall, nonlocalized change in the illumination of the experimental cubicle.

Time intervals between signals were controlled by two Hunter electronic timers. The signal would not occur unless *S* was holding down the rest button. Time from the onset of the signal to release of the rest key (i.e., reaction time), and time from release of the rest key to the pressing of the appropriate response key (i.e., movement time) were recorded on Hunter klockkounters. Recording of data and equipment operation were done silently, and incidental noise was masked by the hum of an electric fan in *S*'s room.

At the conclusion of the experimental procedure *S* was paid in accordance with a crude estimate of his performance and cautioned not to discuss the experiment with other students.

RESULTS

The unit of response used in the statistical analyses was speed of response, that is, the reciprocal of reaction and movement time in seconds. Speed scores were used on the assumption that they are linearly related to reaction potential (Spence, 1956, p. 117). In addition, this transformation tended to reduce the influence of excessively long response times.

Both within- and between-groups analyses of variance were performed on these data. The movement-time measures (i.e., time from release of the rest key to the pressing of the appropriate response key) were found to be unreliable and not related to the independent variables in any systematic way. Apparently the effects of generalized stimulus and temporal cues were manifested before *S* released the rest key and the movement-time scores were mainly a function of random error and key distance, which was held constant. Therefore, only the analyses of the reaction-time measures (i.e., time from signal onset to release of the rest key) are considered in the remainder of the presentation.

Within-Groups Comparisons

Group C.—The strength of the tendency to respond to the positive and negative stimuli at each of the test intervals is shown in the left panel of Fig. 1. Speed of response to both kinds of stimuli increased with nearness in time to the 12-sec. interval; i.e., gradients are evident. It is also evident that the negative gradient was higher and, in confirmation of the theoretical prediction, steeper than the positive gradient. The analyses of variance of the within-groups comparisons are summarized in Table 1 (see the Group C summary). As is indicated in Fig. 1, temporal gradients were established (I effect), and the shock gradient was higher than the

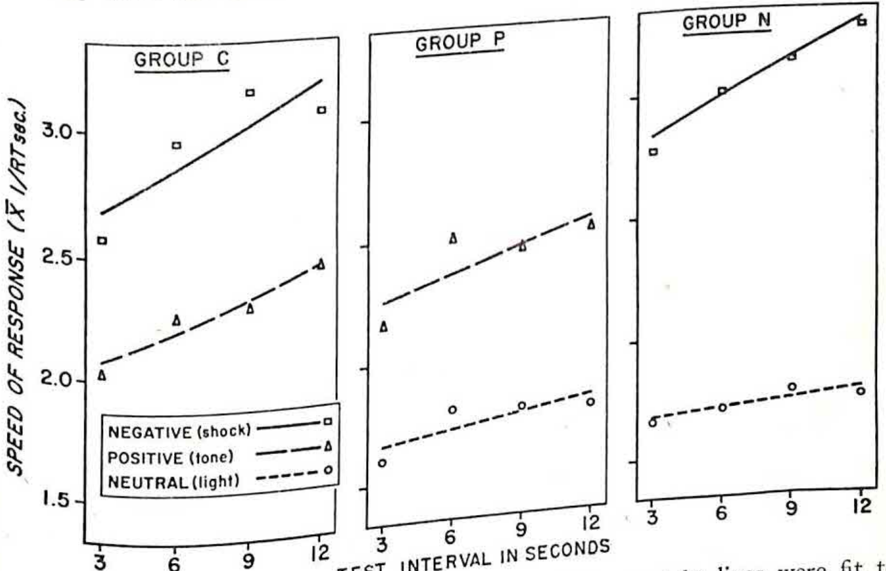


FIG 1. Temporal gradients of response strength. (The straight lines were fit to the data points by the method of least squares. Each point is the mean of speed of response scores for 30 *Ss*.)

TABLE 1

WITHIN-GROUPS ANALYSES OF VARIANCE

Source	*df*	Group C		Group P		Group N	
		MS	*F*	*MS*	*F*	*MS*	*F*
Subjects (*S*s)	29						
Stimulus (S)	1	54.851	39.94***	52.821	119.69***	220.458	107.20***
Interval (I)	3	5.093	26.51***	1.789	13.08***	2.043	12.27***
Half (H)	1	2.190	21.79***	1.028	8.40**	1.038	4.63*
S×I	3	.569	3.21*	.140	1.18	.988	6.69***
S×H	1	.010	<1	.060	<1	.056	[illegible]
I×H	3	.238	1.44	.209	1.64	.335	2.22
S×I×H	3	.736	4.47**	.262	2.57	.504	3.36*
*S*s×S	29	1.373		.441		2.057	
*S*s×I	87	.192		.137		.166	
*S*s×H	29	.100		.122		.224	
*S*s×S×I	87	.177		.119		.148	
*S*s×S×H	29	.164		.079		.172	
*S*s×I×H	87	.165		.127		.151	
*S*s×S×I×H	87	.165		.102		.150	
Pooled error	435	.249		.140		.287	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

tone gradient (S effect). The significant H effect was due to an increase in speed of response from the first half of the test trials to the second half. It is probable that this effect in this and the other analyses merely reflects learning during the testing situation.

The Group C analysis further showed that the S × I interaction, the interaction implying slope differences, was significant. Separate analyses of the trends of the negative and positive means revealed significant linear components for each at the .001 level. It may be asserted then, that the linear components of each of the gradients contributed significantly to the obtained interaction and that, therefore, the gradients differed in steepness. From inspection of Fig. 1, it may be seen that this difference was in the direction predicted from conflict theory. The analysis also indicated that the S × I × H interaction, i.e., the change in the S × I interaction from the first half to the second half, was significant. The slope differences seen in Fig. 1 were present only in the first half of the test trials. The differences between halves, however, were not sufficient to eliminate significance of the overall S × I interaction.

Group P.—The center panel of Fig. 1 shows mean speed of response to the positive and neutral stimuli for each of the four test intervals. The summary of the Group P analysis, Table 1, shows the positive gradient was higher (S effect), but not steeper than the neutral gradient. The gradient (I) effect was significant, and separate tests of the linear components of the gradients revealed significance (.01 level).

Group N.—The functions relating response speed to test-interval length for the negative and neutral stimuli are shown in the right panel of Fig. 1. Gradients were established (I effect), and the negative gradient was higher (S effect) and steeper (S × I interaction) than the neutral gradient (see the summary of the Group N analysis, Table 1). Separate tests of the linear components of the gradients reached significance (.001 level for the negative gradient and .05 level for the neutral gradient). The significant S × I × H interaction was due principally to changes in nonlinear components of the gradients for each half and, therefore, no interpretation was suggested by the data.

Between-Groups Comparisons

Groups C and N (negative gradients).—Visual inspection of the negative gradients shown in Fig. 1 suggests that the negative gradient established in Group C was probably not different from the negative gradient in Group N. The analyses of variance of the combined shock measures showed no significant differences due to height (Group treatment) or to slope (Group × Interval interaction). As one would expect from the separate within-groups analyses, there was a significant negative gradient effect (.001 level for the Interval treatment) and a significant learning effect across halves (.05 level). The Interval × Half and Group × Interval × Half interactions reached significance (.05 and .01 level, respectively) and are seen as largely due to changes in the nonlinear components of the separate gradients for each half.

Groups C and P (positive gradients).—Inspection of the positive gradients shown in Fig. 1 suggests that these gradients did not differ in height or slope. The analysis of variance did not show height (Group) or steepness (Group × Interval) differences; it may therefore be assumed that the positive gradient did not vary between Groups C and P. The Interval × Half interaction reached significance (.05 level) but this was again due mainly to changes in nonlinear components across halves.

Groups P and N (neutral gradients).—Examination of the neutral gradients shown in Fig. 1 shows that, although neutral gradients were established (.001 level for the Interval treatment), neither height nor steepness differences are evident. None of the interaction effects reached significance.

Group C (positive gradient) and Group N (neutral gradient).—The measures comprising the positive gradient from Group C and the neutral gradient from Group N were compared in this analysis. As might be expected from visual inspection of these gradients in Fig. 1, the analysis revealed that the tone gradient (right panel) was higher (Group effect, .001 level) and steeper (Group × Interval effect, .01 level) than the light gradient (left panel). Again, significant interval and half-treatment effects reflected gradient and learning effects, respectively.

This between-groups analysis was performed in light of the nonsignificant S × I interaction in the within Group P comparisons. The reader should note that although there was no evidence of slope difference between the gradients from Group P, the tone and light gradients in Groups C and N did differ in steepness. There is some basis, then, for assuming that the monetary reinforcement following the response to the tone signals had an incremental effect on the slope and height of the positive gradient, an effect which was absent from the gradient of response speed to the unrewarded light stimuli.

DISCUSSION

The most noteworthy feature of the results is the establishment of positive (approach) and negative (avoidance) gradients which differ in steepness in the direction predicted from conflict theory. In the context of the type of theory generated by Miller's work, the gradients of increasing response tendency may be attributed to a temporal gradient of generalized habit strength established through the operation of the principle of stimulus generalization. The habit-strength gradients for both the positive and negative stimuli may be identical, if the internal and external cues characteristic of the 12-sec. interval are identical. If the underlying *gradient* of habit strength is multiplied by a large drive value in the case of the shock stimulus and a smaller drive value produced by the tone, the overt negative gradient would be higher and steeper than the positive gradient. This supposition concerning drive would seem justified on the basis that the drive value of the shock may result from pain, fear of pain, and fear of loss of money while the tone drive value results primarily from the wish for money. Whether the respective drive values are seen as constant at all temporal intervals or as gradient functions it is unlikely that the positive drive factors would outweigh the negative factors in this situation.

It is doubtful that the positive- and negative-gradient differences are attributable merely to differences in the sensory modalities of the stimuli. Tone and shock signals were used because with respect to reaction-time data they probably do not differ (Woodworth & Schlosberg, 1955, p. 26). In addition, the test stimuli were presented with reduced intensity to attenuate the influence of dynamogenic factors, and a crude attempt to equate these properties was made by E in selecting the test intensities used. There is evidence, however, that simple reaction times to light are slower than to tone or shock. For this reason conclusions based on the gradients generated by the signals used for counterbalancing purposes (lights) must be viewed as tenuous.

To account for gradients that cross, a result which might be obtained in temporal conflict situations, one might include the supposition of crossing habit-strength gradients or of intersecting gradients of drive. It is also possible in this theoretical framework to predict gradients which would not intersect, but would coincide. If this were the case, the conflict condition might be far more disruptive to S than if the gradients crossed in the usual way. But, without specific evidence of crossing or coincident temporal gradients, further elaboration of these speculations would seem premature.

For the present, a more appropriate concern is whether one may conclude greater negative gradient slope in the absence of empirical intersection of the gradients. Miller has suggested that if the obtained gradients cross, then the slope differences cannot be attributed to any distorted relationship between the overt response and the underlying gradient of excitatory potential (Miller, 1959, p. 222; Miller & Murray, 1952, p. 229). This is perfectly true, but it might lead one to assume erroneously that without empirical intersection there is no way of predicting whether obtained slope differences are artifactitious or reflect excitatory-potential gradients which do differ in steepness.

For example, if strength of response is a linear function of excitatory potential, which assumption seems reasonable when speed measures are used, the empirical slope differences must reflect slope difference between the underlying gradients. If overt response strength is a positively accelerated function of excitatory potential then distortions can occur, but this relationship is doubtful considering physiological limits to response strength. The relationship might be negatively accelerated, i.e., the increment in strength of response per unit of increase in excitatory potential would become smaller as the potential increased in strength. In this instance if the *higher* of the two empirical gradients is found to be steeper this must result from larger increments in excitatory potential than those which underlie the flatter, less elevated gradient. Therefore, if a negatively accelerated function may be assumed and if the more elevated response-strength gradient is found to be steeper, then the greater slope must follow from a steeper gradient of excitatory potential and not from a distorted relationship between the physical units of speed and the psychological units of excitatory potential. If a negatively accelerated relationship holds or if the function is linear, it may be concluded that the obtained negative gradient is steeper than the positive gradient even in the absence of empirical intersection.

REFERENCES

Brown, J. S. A note on a temporal gradient of reinforcement. *J. exp. Psychol.*, 1939, **25**, 221–227.

Brown, J. S. Principles of intrapersonal conflict. *J. conflict Resolut.*, 1957, **1**, 135–154.

Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944.

Miller, N. E. Extensions of liberalized S-R theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

Miller, N. E., & Murray, E. J. Displacement and conflict: Learnable drive as a basis for the steeper gradient of avoidance than of approach. *J. exp. Psychol.*, 1952, **43**, 227–231.

Mowrer, O. H. Preparatory set (expectancy): Some methods of measurement. *Psychol. Monogr.*, 1940, **52**(2, Whole No. 233).

Rigby, W. K. Approach and avoidance gradients and conflict behavior in a predominantly temporal situation. *J. comp. physiol. Psychol.*, 1954, **47**, 83–89.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, 1955.

(Received April 10, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 242–249

DISCRIMINATION LEARNING AS A FUNCTION OF PRIOR RELEVANCE OF A PARTIALLY REINFORCED DIMENSION[1]

FRED ABRAHAM,[2] ISIDORE GORMEZANO, AND RICHARD WIEHE

Indiana University

This experiment attempted to assess the relationships between learning to sort cards on the basis of a relevant dimension after a nonreversal shift, the amount of simultaneous partial reinforcement of another dimension, and the prior relevance or irrelevance of that partially reinforced dimension. A $2 \times 2 \times 4$ factorial design analysis with 160 human Ss indicated that learning was related to the amount of partial reinforcement as measured by the number of various types of responses to criterion ($p < .001$ for some of these) and was unrelated to the prior relevance of the partially reinforced dimension. The former is a replication of earlier reports, while the latter parsimoniously dictates restrictions upon the interpretation of the relationships between partial reinforcement and other variables in discrimination studies.

The purpose of this experiment was to assess the relationship between prior relevance or irrelevance of the color or form dimension and number sorting in the Wisconsin Card Sorting Test (Grant & Berg, 1948) when color or form was intermittently reinforced 0, 25, 50, or 75% of the time during number sorting.

In a study by Buss (1953), intermittent reinforcement of a previously relevant dimension was inadvertently involved as a confounded variable in the comparison of reversal shifts (where the relevant dimension stays the same but the positive and negative stimuli are reversed) vs. nonreversal shifts (where a formerly irrelevant dimension becomes relevant). Buss attributed the observed inferiority of nonreversal learning, in part, to intermittent reinforcement of a previously relevant dimension which tended to support a perseverative response tendency. Subsequent investigations (Buss, 1956; Harrow & Friedman, 1958; Kendler & D'Amato, 1955) have substantiated this inferiority of nonreversal learning, even when intermittent reinforcement of a previously relevant dimension was eliminated (Harrow & Friedman, 1958). However, none have systematically studied the relationships between amount of intermittent reinforcement and prior relevance or irrelevance of the intermittently reinforced dimension.

While a few studies (Buss, 1959; Gormezano & Abraham, 1961a) have shown an increase in difficulty with increasing amounts of intermittent reinforcement of a previously relevant dimension following nonreversal shifts, it has not been demonstrated that prior relevance is a necessary condition for this increase in negative transfer. Studies by Gormezano and his associates (Abraham & Gormezano, 1960; Gormezano & Abraham, 1961b; Gormezano & Grant, 1958) suggest, in fact, that an increase in difficulty with increasing amounts of intermittent reinforcement of another dimension, along with the relevant dimension, occurs regardless of the prior reinforcement history of that partially reinforced dimension. The present investigation, then, attempts to assess the relationship of both variables, the amount of intermittent reinforcement and the prior reinforcement history, to learning following a nonreversal shift.

---

[1] This investigation was carried out while F. Abraham held a United States Public Health Service Predoctoral Research Fellowship, and with the aid of a National Science Foundation grant to I. Gormezano. A brief report was presented at the APA meetings in Chicago, 1960.

[2] Now a fellow of the USPHS Interdisciplinary Research Training Program, Brain Research Institute, University of California at Los Angeles Center for the Health Sciences.

METHOD

Apparatus and general procedure.—The apparatus and general procedure are similar to the modifications of the WCST (Grant & Berg, 1948) developed in the studies of Gormezano and Grant (1958) and Gormezano and Abraham (1961a, 1961b). Cards were used which contained one to four identical figures (triangles, stars, crosses, or circles) of a single color (red, green, yellow, or blue). In the test situation, a gray sorting tray was placed on a table before *S*. Four stimulus cards consisting of one red triangle, two green stars, three yellow crosses, and four blue circles were placed from right to left in the upper halves of the four double compartments. Several decks of 96 response cards were composed from certain cards selected from the 64 possible combinations of the four values on each of the three dimensions (number, figure, color). The cards used for this function were of two types, ambiguous and unambiguous. Unambiguous cards were selected such that they would match each of three of the four stimulus cards on one value of each of the three dimensions (e.g., two red crosses). Ambiguous cards were selected such that each matched a stimulus card on the basis of the number of figures (relevant dimension) while simultaneously matching on the basis of one of the other dimensions (partial dimension). Such a card would match one of the other stimulus cards on the basis of the remaining dimension (irrelevant dimension), and not match the remaining two stimulus cards on any of the dimensions (e.g., one red star). Six decks (ambiguous decks) were composed of both ambiguous and unambiguous cards such that two decks contained 24 (25%) ambiguous cards, two contained 48 (50%), and two contained 72 (75%); one deck at each percentage contained ambiguous cards that matched on the basis of color simultaneously with number; the other deck on the basis of form simultaneously with number. Another deck (unambiguous deck), replicated several times because of its greater use, was composed of only unambiguous cards (0% ambiguity).

The *E* and *S* sat in a small room facing each other across a small table upon which the sorting board was placed. The three decks selected for use were placed in a partitioned box fastened to one of the legs of the table, facing *E* and out of view of *S*. The *S* was instructed as follows:

> I am going to present you with a series of cards similar to the ones you see before you. Every time I give you a card, I want you to place it in the compartment underneath the card to which you think it belongs. Each time you place a card I will tell you whether you are right or wrong. After a while I will stop saying right or wrong, but you are to continue placing the cards.

The response cards were then presented to *S*, one at a time in a standardized manner and order. Initially, cards from the 0% unambiguous deck were used and sorts that matched on the basis of one dimension, color or form, were reinforced (Stage I). After 10 consecutive correct sorts, *E* then shifted to another deck, either ambiguous or unambiguous, and reinforced only *S*'s number sorts (Stage II). After 10 consecutive correct number sorts, 10 unambiguous cards were sorted without further reinforcement (Stage III), and the experimental session then terminated. If *S* did not reach the criterion in either Stage I or II, his data were omitted from the analysis.

Subjects and experimental design.—The *S*s were 173 men and women from Indiana University elementary psychology courses: 10 in each of 16 groups and 13 failures. Of the 13 failures, 1 failed to meet criterion in Stage I; the other 12 failed to meet criterion in Stage II. The Stage II failures were unrelated to any of the experimental variables, as tested by separate chi-square goodness-of-fit tests for each of the variables.

Half of the groups were reinforced for color sorts in Stage I, the other half for form (the "previously" or "prior relevant dimensions"). All groups received 100% reinforcement for number sorts (the "relevant dimension") in Stage II, and by choice of the appropriate deck, 0% (with an unambiguous deck), 25%, 50%, or 75% reinforcement of another dimension, color or form (the "partial dimension") on these sorts (with ambiguous decks), and 0% reinforcement of the remaining, third dimension (the "irrelevant dimension"). For eight of these groups, the partial dimension was the same as the previously relevant dimension, one group at each percentage level for form as the partial dimension and one at each percentage level for color; for the other eight groups, the partial dimension was the dimension other than the previously relevant dimension, one at each percentage level for each partial dimension, form and color. Thus the design was a 2 × 2 × 4 factorial with two relevant dimensions, form and color, in Stage I; with two conditions of similarity between the Stage I dimension and the Stage II partial dimension (i.e., color relevant in Stage I, number relevant with color partial in Stage II, or similarly form to number with form) or different (i.e., color to number with form, or form to number with color); and with four amounts of partial reinforcement of the Stage II partial dimension, 0, 25, 50, or 75%.

RESULTS AND DISCUSSION

Stage I.—In Stage I, the total number of trials to reach criterion for each *S* was transformed by taking the square root of the score plus one half. The mean of the transformed scores for all *S*s was 1.646. Most of these responses were incorrect; the mean of the transformed error scores being 1.570 and the mean of the transformed correct scores being .850. The variability in the total trials to criterion was unrelated to the variables defined by Stage II procedures, but was related to the Stage I dimension, $F\,(1, 144) = 25.70$, $p < .001$, with color being the harder dimension to be learned, the mean being 2.135 compared to 1.156 for the form *S*s, a finding which replicated an earlier report by Grant and Curran (1952).

Stage II.—As in Stage I, the 10 criterion trials were excluded from the response measures. The square-root transformation was performed on the following measures for each *S*: total number of trials, number of incorrect responses, number of correct responses, number of perseverative errors, number of nonperseverative errors, number of ambiguously correct responses, and number of unambiguously correct responses. Analyses of variance were computed for the total trials, total errors, and total correct (Table 1). Total errors do not increase with increasing partial reinforcement (Fig. 1B), so that any increase in total trials (Fig. 1A) must be due to the increase in correct responses (Fig. 1C). Errors could be perseverative errors (responses for

TABLE 1

ANALYSES OF VARIANCE FOR VARIOUS SCORES MADE WITH ALL RESPONSE CARDS IN STAGE II

Source	*df*	*F*		
		Total Trials	Total Errors	Total Correct
Color or form (Stage I) (C-F)	1	2.95	3.32	3.70
Same or different (S-D)	1	.52	1.79	.08
% reinforcement	3	3.49*	1.59	16.55***
C-F × S-D	1	.09	.01	.25
C-F × %	3	2.63	4.13**	1.96
S-D × %	3	.35	.69	.42
C-F × S-D × %	3	.90	1.11	1.14
Within-groups *MS*	144	(3.47)	(1.36)	(1.59)

* $p < .05$.
** $p < .01$.
*** $p < .001$.

which the response card did not match the stimulus card on the basis of the dimension relevant in Stage II but did match on the basis of the dimension relevant in Stage I), nonperseverative errors (responses for which the response card did not match the stimulus card on the basis of either the relevant dimension in Stage I or the relevant dimension in Stage II but did match on the basis of the remaining dimension), and unique response errors (responses for which the response card did not match the stimulus card on the basis of any of the principal dimensions of color, number, or form). As with total errors, both perseverative and nonperseverative errors vary independently of any of the independent variables (Fig. 2A and 2B). Unique responses were too few and too randomly distributed in all three stages to warrant further analysis. Correct sorts can be differentiated on the basis of whether they were made with ambiguous or unambiguous response cards. Most of the correct responses were made with the ambiguous cards, there being extremely few correct responses made with unambiguous cards (Fig. 2C and 2D).

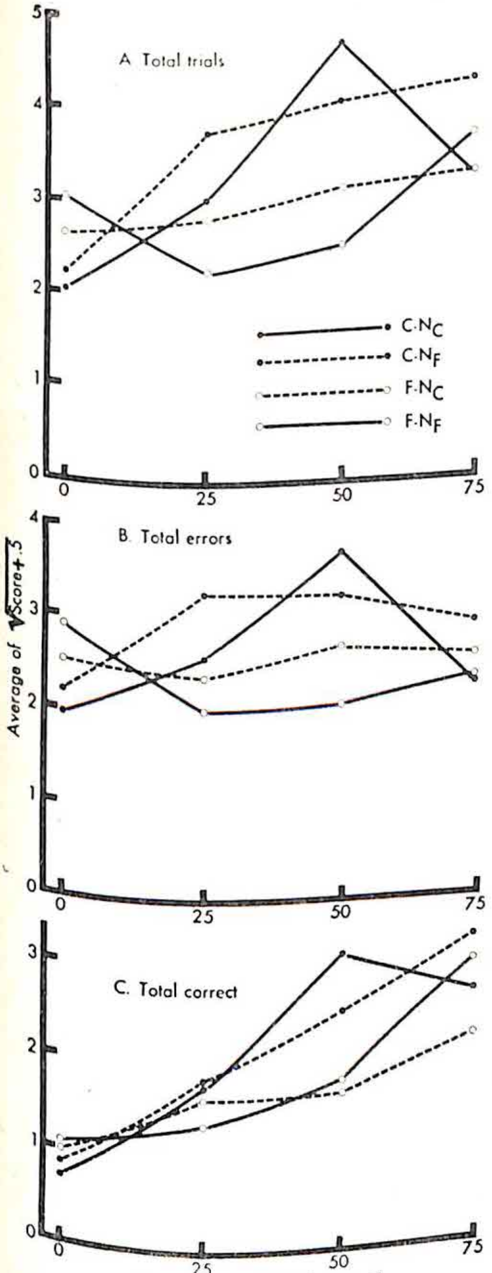


FIG. 1. Stage II scores for all response cards (exclusive of criterion series) for all groups as a function of amount of partial reinforcement with the Stage I relevant dimension and the similarity of the Stage II partial dimension to the Stage I relevant dimension as the parameters.

These measures clearly indicate that *Ss* responded correctly with ambiguous cards (which match a stimulus card on two dimensions rather than just one), independently of whether the partial dimension was the same as the previously relevant dimension or not. There are more perseverative than nonperseverative errors, but these measures are unrelated to the amount of partial reinforcement or the Stage I relevance of the Stage II partial dimension. The Gormezano and Grant (1958) experiment, where the partial dimension is the same as the prior irrelevant dimension, shows essentially the same results, except that perseverative and nonperseverative errors were equal. The Gormezano and Abraham (1961a) and Buss (1959) studies indicated an increase in difficulty of learning following a nonreversal shift with increasing amounts of partial reinforcement of a previously relevant di-

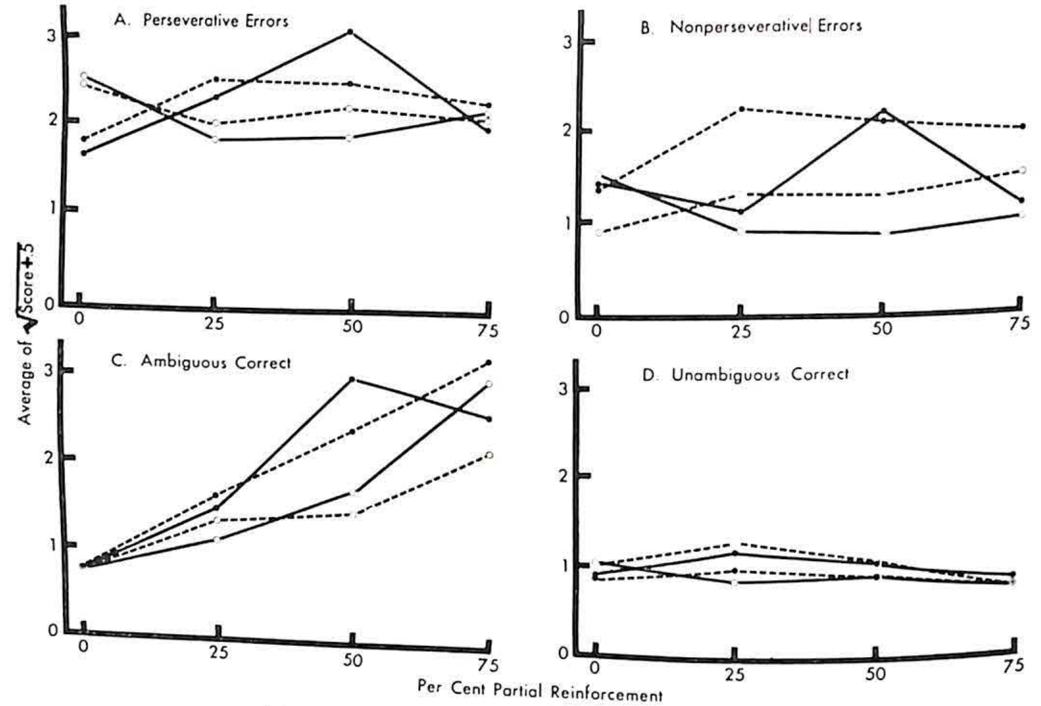


FIG. 2. Stage II scores for all response cards.

mension. While their results are not presented in sufficient detail, they are probably attributable principally to the tendency to respond correctly on cards that are simultaneously correct in Stage II on the basis of two dimensions at once—the present and the previously relevant dimensions. Because of this tendency to respond correctly with ambiguous cards, the relationship of amount of partial reinforcement and prior reinforcement history to Stage II learning is best assessed with analysis of responses made with unambiguous cards.

Therefore, further analyses were based on responses made only with the unambiguous response cards. The percentage of these cards sorted perseveratively incorrectly, nonpersever-

TABLE 2

ANALYSES OF VARIANCE FOR VARIOUS SCORES MADE WITH UNAMBIGUOUS RESPONSE CARDS IN STAGE II

Source	*df*	*F*		
		Perseverative Errors	Nonperseverative Errors	Correct
Color or form (Stage I) (C-F)	1	13.62***	21.33***	.08
Same or different (S-D)	1	40.41***	538.05***	.13
% reinforcement	3	.59	.19	2.00
C-F × S-D	1	2.00	3.03	.19
C-F × %	3	.70	1.11	.38
S-D × %	3	9.55***	14.57***	.19
C-F × S-D × %	3	.24	.16	.43
Within-groups *MS*	144	(381.98)	(276.84)	(215.30)

*** $p < .001$.

atively incorrectly, and correctly by each *S* was transformed to degrees by means of an arc-sine transformation. Analyses of variance were computed for each of these three measures (Table 2). Perseverative errors (*a*) were greater for the groups having the easier Stage I dimension, form, (*b*) were greater for the groups with the partial dimension the same as the Stage I dimension, (*c*) increased with increasing percentage of partial reinforcement for those groups with the partial dimension the same as the Stage I dimension, and (*d*) decreased with increasing percentage of partial reinforcement for the groups with the partial dimension different from the relevant dimension in Stage I (Fig. 3A). Nonperseverative errors (*a*) were greater for the groups having the more difficult Stage I dimension, color, (*b*) were greater for the groups with the partial dimension different from the Stage I dimension, (*c*) increased with increasing percentage of partial reinforcement for the groups with the partial dimension different from the Stage I dimension, and (*d*) decreased with increasing percentage of partial reinforcement for the groups with the partial dimension the same as the relevant dimension in Stage I (Fig. 3B). Perseverative errors ($\bar{\phi} = 63.73$) were greater than nonperseverative errors ($\bar{\phi} = 20.44$). Extremely few correct responses were made with these unambiguous cards ($\bar{\phi} = 6.79$).

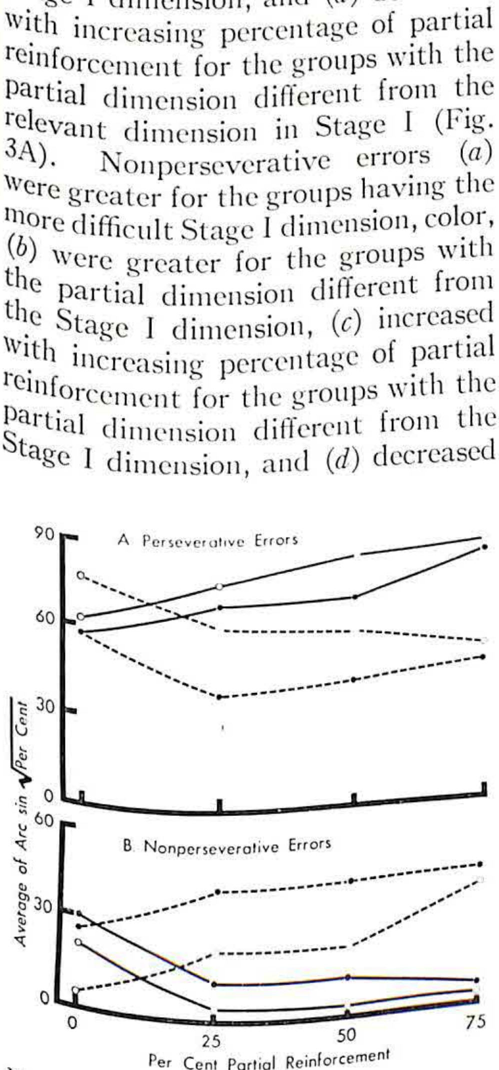


FIG. 3. Stage II scores with unambiguous cards only.

It appears, then, that increased percentage of reinforcement of the partial dimension does not increase the probability of an error, but does increase the probability that an error, when made, will be made on the basis of the partial dimension, independently of whether or not the partial dimension, was the same as the previously relevant dimension (note in Table 2 the significant Same-Different × Partial Reinforcement sources of variance for the perseverative and nonperseverative errors). The probability that an error with an unambiguous card, when made, will be a perserverative one is greater than that it will be a nonperseverative one, but this difference is not a function of the amount of partial reinforcement (note in Table 2 the small percent reinforcement sources of variance for the perseverative and nonperseverative errors made with unambiguous cards). That is, it appears that the probability of a correct or incorrect response with an unambiguous card does not change following correct responses with ambiguous cards, but that the probability of an error made on the basis of the partial dimension increases with increasing amounts of such partial reinforcement, while the probability of an error made on the basis of the irrelevant dimension correspondingly decreases.

The fact that the probability of a correct response with unambiguous cards does not increase the greater the number of ambiguously correct trials is more likely a limitation of this study rather than the principle of reinforcement. There was little variability in the unambiguously correct scores, there being few errors made following the first such response. Also, direct measures of the relevant probabilities, such as the conditional probability of a correct response with an unambiguous card given zero, one, two, or three ambiguous cards, correctly sorted, on the immediately preceding trials, were not made. Likewise, the lack of a significant relationship between amount of partial reinforcement and the prior relevance of the partial dimension could be a limitation of the sensitivity of the present study, rather than a limitation of the principle of reinforcement. Hunt (1962) has interpreted such a lack of interaction between Stage I and Stage II variables (e.g., as in the 1958 Gormezano and Grant study where no relation was found between Stage II learning and amount of partial reinforcement in Stage I of the Stage II relevant dimension) as support "for a mediating (either verbal or orienting) response in S-R analysis of concept learning." Despite efficient designs, however, analysis of variance tests of trial to criterion measures could be rather insensitive to small differences in exact probability measures that would be indicative of such interactions. Until designs and analyses with improved sensitivity are performed, the contribution of these studies to the need for such complicating axioms to behavior theory might best be parsimoniously viewed with reservation.

The present experiment does clearly indicate a relationship between learning to respond on the basis of one dimension and the amount of partial reinforcement that another dimension receives simultaneously. This relationship is manifest partly in the increasing relation between total correct responses to criterion and amount of partial reinforcement because of the tendency to sort ambiguous cards correctly, and is especially manifest in the increasing relationship between the proportion of responses with unambiguous cards made on the basis of the partial dimension and the amount of partial reinforcement. This study indicates that these relationships occur both when the partial dimension was previously relevant and when it was previously irrelevant. Gormezano and Grant (1958), Gormezano and Abraham (1961a), and Buss (1959) have similarly shown increased difficulty of learning with increased partial reinforcement of the previously relevant condition only. Gormezano and Grant (1958) and Gormezano and Abraham (1961b) have also found this relationship in Stage I where the partial and other dimensions had no prior reinforcement history, and the Gormezano and Abraham (1961b) study has demonstrated it for the case where the partial dimension was a previously partial dimension. Thus the variations in amount of partial reinforcement of a partial dimension seems to be the major controlled source of variance in discrimination learning where such partial reinforcement exists, independently of the prior reinforcement history of the partial dimension. Hunt (1962) has been aware of this possibility (see Ch. 3 of his scholarly treatise *Concept Learning*, especially Footnote 7).

Stage III.—In Stage III, *S*s responded with number sorts on all but 101 of 1,600 trials, the mean of the square-root transformed scores of all other responses being .929. These other responses increased slightly over trials, but were not related to any of the independent variables (for all *F*s, $p > .001$), and were equally likely to have been color or form sorts independently of the independent variables. Thus it would seem that the 10-trial criterion in Stage II was sufficient even for the 75% groups. This high resistance to extinction is similar to that often found with verbal operants (Buchwald, 1960).

REFERENCES

ABRAHAM, F. D., & GORMEZANO, I. Concept formation as a function of prior relevance of a partially reinforced dimension. *Amer. Psychologist*, 1960, **15**, 467. (Abstract)

BUCHWALD, A. M. Supplementary report: Alteration in the reinforcement value of a positive reinforcer. *J. exp. Psychol.*, 1960, **60**, 416–417.

Buss, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 75–81.

Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.

Buss, A. H. Supplementary report: Maintenance of a previously learned concept as a function of partial reinforcement. *J. exp. Psychol.*, 1959, **58**, 414–415.

Gormezano, I., & Abraham, F. D. Intermittent reinforcement of a formerly relevant dimension in concept formation. *Psychol. Rep.*, 1961, **8**, 111–116. (a)

Gormezano, I., & Abraham, F. D. Intermittent reinforcement, nonreversal shifts, and neutralizing in concept formation. *J. exp. Psychol.*, 1961, **61**, 1–6. (b)

Gormezano, I., & Grant, D. A. Progressive ambiguity in the attainment of concepts on the Wisconsin Card Sorting Test. *J. exp. Psychol.*, 1958, **55**, 621–627.

Grant, D. A., & Berg, E. A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigle-type card-sorting problem. *J. exp. Psychol.*, 1948, **38**, 404–411.

Grant, D. A., & Curran, J. F. Relative difficulty of number, form and color concepts of a Weigel-type problem using unsystematic number cards. *J. exp. Psychol.*, 1952, **43**, 408–413.

Harrow, M., & Friedman, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, **55**, 592–598.

Hunt, E. B. *Concept learning: An information processing problem.* New York: Wiley, 1962.

Kendler, H. H., & D'Amato, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

(Received April 18, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 250–255

CONJUNCTIVE AND DISJUNCTIVE CONCEPT FORMATION UNDER EQUAL-INFORMATION CONDITIONS[1]

MICHAEL B. CONANT AND TOM TRABASSO[2]

Stanford University

College students learned sets of conjunctive and inclusive disjunctive concepts in which the minimum number of choices necessary for solution was equated. The disjunctive concepts were more difficult to master. Selection of instances and redundancy in choices indicated that Ss learn to use a positive-focusing strategy within a conjunctive problem sooner than they learn to use a negative-focusing strategy within a disjunctive problem.

In a discussion of the logical structure of concepts, Bruner, Goodnow, and Austin (1956) made particular note of the difference between conjunctive and disjunctive concepts. A conjunctive concept is defined by the joint presence of several values. An example of a conjunctive concept is "blue-eyed *and* red-haired men." A disjunctive concept, on the other hand, is defined by the presence of one or another value (e.g., "tall *or* thin men"). Hunt (1962) makes a further distinction between inclusive and exclusive disjunctivity. In an inclusive disjunction, either value or both may occur; in an exclusive disjunction, either value but not both may occur.

In studies of concept formation, Bruner et al. found that *S*s have difficulty working with disjunctive concepts; the authors suggested that this was due, in part, to the necessity of using negative instances in order to solve disjunctive problems most efficiently. Indirect evidence for this hypothesis came from an earlier study by Hovland and Weiss (1953), who found that *S*s are inefficient in using negative instances to solve conjunctive problems. Hunt and Hovland (1960) studied problems in which conjunctive, relational (e.g., larger than), and disjunctive solutions were equally correct. They found conjunctive and relational solutions are more frequently offered than disjunctive ones. These authors suggest that the frequency of each type of solution may be an indication of its relative difficulty. Recently, Wells (1963) obtained transfer of training from disjunctive concept formation to choice problems where either a conjunctive or a disjunctive solution was possible, indicating that prior learning of concept types is a factor in the difficulty of concept learning. Wells' training procedure apparently modified prior habits of focusing on positive instances and thereby facilitated disjunctive concept formation.

The purpose of the present study was to obtain a direct comparison between conjunctive and disjunctive problems given to the same *S*s. This aim was accomplished by constructing conjunctive- and disjunctive-concept problem sets which were structurally, perceptually, and informationally equivalent.

---

[1] Acknowledgment is made to David L. Brown for his guidance during the execution of the research. The second author was supported by a postdoctoral research fellowship, No. 5 F2 MH-18,070-02, from the National Institute of Mental Health, United States Public Health Service.

[2] Now at the University of California, Los Angeles.

METHOD

Stimuli.—The stimuli consisted of two colored figures drawn on white 3 × 5 in. file cards. On each card there was a triangle on the left and a circle on the right. The figures varied independently in two dimensions, size (large or small) and color (red or green), making 16 different patterns.

Procedure.—The 16 cards used throughout the experiments were placed in a 4 × 4 array before *S* at the start of each problem. The dimensions and values were described to *S*, and appropriate instructions were read before each conjunctive and disjunctive set of problems; it was ascertained that *S* knew the general nature of the relevant concept type by having him pick out all positive and negative example cards associated with a sample concept of each type. To begin a problem, *E* presented *S* with 1 of the 16 cards as either a positive or negative example of an unknown concept. The *S* then selected single cards and was told each time whether his choice was a positive or negative instance. The *S* grouped the selected cards according to class, and a problem was completed when he verbalized the concept. Premature verbal solutions were discouraged.

At the opening of a session, the following instructions were read:

> This is an experiment to see how you think. As you can see, there are 16 cards laid out on the table. Each card contains one circle and one triangle. Circles and triangles vary both in size and color. For instance, a circle is large or small and red or green. The same holds for triangles.
>
> For purposes of this experiment, a concept will be considered to be a certain set of these cards. A concept about these cards has been chosen. Your job will be to find out the concept as efficiently as possible, in a manner that will be described to you.

Then, before a set of conjunctive problems was given to *S*, the following three paragraphs were read:

> An example of the type of concept we are dealing with might be all those cards which contain *both* large triangles and green circles. Would you please point out all the cards here which have the property of having both large triangles and green circles?
>
> That is correct. Four of the 16 cards have this property, and with the type of concepts we will now deal with, 4 of the cards will always be examples of the concept, and 12 will not.
>
> Please keep in mind that with the type of concepts we are now dealing with, just one property of the circles is required (red, large, green, or small), and one property (red, large, green, or small), not necessarily the same, is required for triangles. In other words, small red triangles is not a concept we are dealing with, because this has two properties of one figure. Each concept, again, requires that its examples have just one property of the circles *and* one property of triangles. Do you have any questions about this?

Before the set of inclusive disjunctive problems, the following three paragraphs were read:

> An example of the type of concept we are dealing with might be all those cards which contain *either* large triangles or green circles or both. Would you please point out all the cards here which have the property of having either large triangles or green circles or both?
>
> That is correct. Twelve of the 16 cards have this property, and with the type of concepts we will now deal with, 12 of the cards will always be examples of the concept, and 4 will not.
>
> Please keep in mind that with the type of concepts we are now dealing with, just one property of circles is required (red, large, green, or small) and one property (red, large, green, or small), not necessarily the same, is required for triangles. In other words, small red triangles is not a part of a concept we are now dealing with, because this has two properties of one figure. Each concept requires that its examples have just one property of circles *or* one property of triangles. Do you have any questions about this?

Following each set of conjunctive or disjunctive instructions, all *S*s were told:

> I will give you a card that is or is not (and you will be told which) an example of the concept. Your job will be to try other cards, one at a time. I will tell you after each choice whether or not these are examples of the concept. You may guess at the concept at any point, but wrong guesses will result in a subtraction from your score. However, there is obviously no penalty for picking particular cards which are not examples of the concept. Your score will be dependent on how few cards

you have to try before you are sure of what the concept is. When you have arrived at the concept, tell me what it is. If you are correct, that problem will be finished and we will go on to another concept. Time is not a factor, only the number of cards you have to try. Work efficiently, but do not hurry. You may take as long as you wish. You may rearrange the cards during trials in any way that may be helpful to you. We will make two rows at the side, one for cards that are examples of the concept, and a row for those that are not. Do you have any questions?

Design.—A 2 × 2 × 2 design was used with the following variables: (*a*) two sets of three problems, a conjunctive set (C) and an inclusive disjunctive set (D); (*b*) two presentation orders, C-D and D-C; and (*c*) sex. Each *S* was given both C and D sets and was randomly assigned to one of the two presentation orders.

Problems.—Table 1 summarizes the two sets of three C and D problems along with the example card given before each problem for Exp. I (for the problems of Exp. II, see below). Given the instructions and the stimuli, there are 16 possible triangle-value and circle-value solutions for each problem. If an example card is positive for a C problem or negative for a D problem, 12 of these possible solutions are eliminated. Similarly, if an example card is negative for a C problem and positive for a D problem, 4 of the possible solutions are eliminated. Thus, positive C instances and negative D instances yield the most information and their respective selection would lead to the most efficient C or D problem solving. When the example card was positive, a C problem could be solved in two card choices, and a D problem in three or four choices, depending upon *S*'s first card choice. When the example card was negative, a D problem required two choices and a C problem, three or four. The C and D sets of three problems each were equated as to the minimum number of choices necessary for solution, 8–10 choices for Exp. I, and 9–12 choices for Exp. II.

Experiments.—Two experiments were performed. In the second experiment, the D problems in Table 1 were made C problems and vice versa. Thus, all C example cards became negative and all D example cards became positive. The second experiment was conducted to assess whether or not the particular value-value pairs contributed to any possible differences observed in Exp. I.

Subjects.—For Exp. I, the *S*s were 24 volunteers, 12 males and 12 females, from psychology classes at Columbia University. For Exp. II, there were 12 volunteers, 6 males and 6 females, from the same student population.

RESULTS AND DISCUSSION

Efficiency was first evaluated by comparing the number of choices required to solve all three problems of a concept set. Table 2 summarizes the analysis of variance for each experiment.

In both experiments, the D set required more choices to solution than the C set. In Exp. I, the mean total choices were 19.04 for the D set and 13.75 for the C set. For Exp. II, these values were 19.00 and 15.05, respectively. No other main effect or

TABLE 1

CONJUNCTIVE AND INCLUSIVE DISJUNCTIVE SETS USED IN EXP. I

Problem	Concept			Example		
	Triangle		Circle	Triangle	Circle	Instance
C1	G	and	L	S, G	L, R	Positive
C2	L	and	R	S, G	S, R	Negative
C3	R	and	S	L, R	L, G	Negative
D1	L	and/or	R	L, G	L, G	Positive
D2	G	and/or	G	L, R	S, R	Negative
D3	S	and/or	L	S, R	S, G	Positive

Note.—The values of the stimuli were G (Green), L (Large), S (Small), and R (Red).

TABLE 2

ANALYSIS OF VARIANCE ON CHOICES TO SOLUTION

Source	Exp. I			Exp. II		
	df	*MS*	*F*	*df*	*MS*	*F*
Between *Ss*	23			11		
Sex (S)	1	6.02	.21	1	15.04	2.00
Order (O)	1	20.02	.72	1	.37	.05
S × O	1	1.69	.06	1	11.81	1.57
Error (b)	20	28.46		8	7.53	
Within *Ss*	24			12		
Concepts (C)	1	336.02	23.40**	1	92.04	7.20*
C × S	1	3.52	.25	1	15.05	1.18
C × O	1	35.02	2.44	1	.21	.02
C × S × O	1	3.69	.26	1	5.21	.41
Error (w)	20	14.36		8	12.81	

* $p < .05$.
** $p < .01$.

interactions were significant and no transfer between concept types was observed. Product-moment correlations between *S*'s number of choices required on each of the two problem sets were not significant. No *S* required all cards to solve a problem; the maximum number selected by an *S* for a D problem was 14 and for a C problem, 11.

The mean number of choices per problem is reported in Table 3. Relative transfer within each set cannot be properly evaluated because of confounding with the class of example cards and inconsistencies in trend.

If *S* selects instances at random throughout a problem, then the proportion of positive instances chosen should be near .25 for C concepts and .75 for D concepts. However, if *S* uses a "positive-focusing strategy" (cf. Bruner et al., 1956) which is efficient for C problems and, conversely, a "negative-focusing strategy" which is efficient for D problems, these respective proportions should be higher or lower. To study this, each *S*'s problem was divided into halves and the number of positive C and negative D instances chosen per half was counted. These Vincentized percentages were nearly equal for the two experiments and were pooled for the summary presented in Table 4.

Summing over all problems, the proportion of positive C choices and of negative D choices was above chance expectations. The first- and second-half comparisons in Table 4 indicate that *S* learns to select, *within a problem*, a positive instance under C conditions more rapidly than a negative instance under D conditions. The *S* would appear to solve C concepts sooner since he learns to choose positive instances within a C problem

TABLE 3

MEAN NUMBER OF INSTANCES CHOSEN PER PROBLEM

Concept	Problems					
	Exp. I			Exp. II		
	1	2	3	1	2	3
C	4.17	5.21	4.37	5.48	5.25	4.32
D	7.00	5.75	6.29	6.91	5.76	6.33

TABLE 4

PROBABILITY OF A POSITIVE C AND A NEGATIVE D CHOICE

Problems	Proportion per Half					
	Positive C			Negative D		
	First	Second	Trials	First	Second	Trials
1	.24	.56	72	.19	.32	117
2	.22	.61	84	.17	.36	92
3	.22	.64	76	.21	.37	107
Total	.22	.62	232	.20	.35	316

more rapidly than negative instances within a D problem. Since the improvement in selection occurs only during the second half of the problem and not over successive problems, there is no apparent transfer of either a positive- or negative-focusing strategy. The tendency to choose more informative instances may result from the information obtained on initial choices. That is, a positive C card may be easier to find once *S* has selected a few cards, be they positive or negative.

An information analysis was performed on the card choices with respect to the number of redundant and nonredundant card selections to solution. A card choice was defined as redundant if it could not eliminate at least one further incorrect solution beyond those already eliminated by the example card, preceding card choices and, if verbalized by *S*, wrong hypotheses. Admittedly, one does not know exactly what solutions were tried by an *S* but this analysis provides indirect evidence for informational use.

According to this analysis, nearly all incorrect solutions were eliminated by card choices. Eighty-seven percent of the problems were solved with all incorrect solutions eliminated; the remainder were solved with three or less incorrect solutions still possible. The average nonredundant choices were nearly equal: for Exp. I, they were 9.04 for C problems and 8.75 for the D set, and for Exp. II, these values were 10.57 for the C set and 11.20 for the D set. However, more redundancy in choice occurred on D problems. For Exp. I, mean redundant choices were 4.71 for the C set and 10.29 for the D set, $F\ (1, 20) = 32.02$, $p < .01$; and for Exp. II, these values were 4.48 for the C set and 7.80 for the D set, $F\ (1, 8) = 7.61$, $p < .05$. No other main effect or interaction was significant.

The relative frequencies with which an *S* chose positive and negative instances on the final card for a problem are summarized in Table 5. The percentage of these choices which were redundant is also reported.

The very high proportion of positive C instances selected as the last card and the above-chance selection of

TABLE 5

PROBABILITY OF POSITIVE AND NEGATIVE FINAL CARDS

Instance	Concept			
	C		D	
	Choices	Redundant	Choices	Redundant
Positive	.85	.66	.64	.84
Negative	.15	.75	.36	.65

negative D instances are consistent with use of positive- and negative-focus strategies within a problem. Last-choice redundant negatives are more frequent for the C set, whereas redundant positives are more frequent for the D set. This is expected since these selections are more likely to be redundant when chosen later in the series.

Since there is no difference, between C and D classifications except for the labels *S* is forced to apply, the form which *S*'s verbal hypothesis must take may be the critical variable. If an *S* is unfamiliar with the language of a D concept, he may be reluctant to proffer a statement until doubly, hence, redundantly certain. To test this possibility, the trial upon which the first hypothesis (correct or incorrect) was offered was compared in the two sets and no significant differences were obtained. In Exp. I, the mean trial of the first hypothesis per C problem was 3.82 and per D problem was 4.28, matched $t\ (23) = 1.28$, $p > .05$; and in Exp. II, the respective values were 4.55 and 4.69, matched $t\ (11) = .32$, $p > .05$. Of the first hypotheses offered, 75% of the Cs and 56% of the Ds were correct.

An alternative interpretation of the redundancy data can be made on the basis of *S*'s choice behavior. With C concepts, a negative choice following a positive is more likely to be redundant than a positive choice following a negative. The reverse is true for D concepts. For the C set, the conditional probability of a negative following a positive was .52; whereas for the D set, the conditional probability of a positive following a negative was .66. Therefore, in C learning, the fewer redundant choices result from both the lower number of choices to solution and the higher incidence of informationally rich positive choices. In D learning, the redundancy results from the larger number of choices and the high incidence of low-information positive choices. An *S* appears to be more efficient in using positive C choices than negative D choices, even though both instance types yield the same high information. In particular, comparing Problems C1 and D2 of Exp. I, where C1 began with a positive example and D2 with a negative, the ratio of redundant to nonredundant choices in D2 was twice that of C1.

REFERENCES

BRUNER, J. S., GOODNOW, J., & AUSTIN, G. *A study of thinking.* New York: Wiley, 1956.

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.

HUNT, E. B. *Concept learning.* New York: Wiley, 1962.

HUNT, E. B., & HOVLAND, C. I. Order of consideration of different types of concepts. *J. exp. Psychol.*, 1960, **59**, 220–225.

WELLS, H. Effects of transfer and problem structure in disjunctive concept formation. *J. exp. Psychol.*, 1963, **65**, 63–69.

(Received April 29, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 256–262

EFFECT OF CONSTANT VERSUS VARIED PAIRING OF SIMULTANEOUS INTENTIONAL- AND INCIDENTAL-LEARNING MATERIALS WITH DIFFERENT RATES AND NUMBERS OF EXPOSURES[1]

MARILYN E. MILLER AND VIRGINIA LAKSO

University of Wisconsin, Milwaukee

224 Ss were instructed to learn a series of 7 nonsense syllables, each enclosed in 1 of 7 different geometric forms. Half the Ss were instructed to learn the forms (Intentional) and half were given no form instructions (Incidental). Either a given syllable appeared in the same form on all trials (Constant), or it appeared in a different form on each trial (Varied). Stimuli were presented for either 2 or 8 sec. for either 1, 3, 5, or 7 trials. Intentional Ss learned significantly more forms and fewer syllables than Incidental Ss. Both longer exposure and increased trials facilitated intentional learning; increased trials facilitated incidental learning only with longer exposure. Changing the form-syllable relationship from trial to trial facilitated learning of forms.

Much of the research concerning incidental learning has been directed toward determining whether the variables affecting intentional learning also affect incidental learning, and whether the effects are similar. Two such variables which have been investigated separately are rate of exposure (Neimark & Saltzman, 1953; Rosenberg, 1959) and number of presentations (Brown, 1954; Postman & Adams, 1958; Saltzman & Atkinson, 1954). The results have been seemingly inconclusive. A direct comparison of previous results is complicated by the fact that different orienting tasks were used. Mechanic (1962) has suggested that most incidental-learning designs could be placed in one of two categories: Type I, in which *S* is not instructed to learn any of the materials; and Type II, in which *S* is instructed to learn some material and is tested on other material which has been presented concurrently. Type I presumably includes those situations in which *S* is required to rate, match, or simply pronounce the material (orienting task), and is then tested on that material. An example of Type II design would be the case in which *S* is instructed to learn a series of nonsense syllables (orienting task), each presented in a geometric form, and is then tested for recognition or recall of the forms (incidental learning). Although the previously cited studies can be classified as either Type I or Type II, there is no unanimity of results regarding rate and number of exposures within types. This may be due to a further complication. The studies investigating rate of exposure used a different number of presentations in each case, and those investigating number of exposures used a different exposure rate in each case.

The present experiment investigated the possibility of an Exposure Rate × Number of Exposures interaction in a Type II design.

It was felt that the manner in which

---

[1] Supported by National Science Foundation Undergraduate Research Grant G-22832, H. L. Madison, Director.

the orienting-task material and the incidental material were paired during presentation might have a differential effect upon the amount of incidental learning; therefore, that variable was also investigated.

In the present experiment, the orienting task consisted of learning a serial list of nonsense syllables, each presented within a geometric form. The *S*s were either told nothing about the forms (Incidental) or were told to learn them while learning the syllables (Intentional). Thus the design investigates the simultaneous intentional learning of verbal material with either intentional or incidental learning of nonverbal material. A given syllable either appeared in the same geometric form on all trials (Constant) or in a different form on each trial (Varied). The syllables (and forms) were presented for either 2 or 8 sec. for a total of either one, three, five, or seven trials.

METHOD

Design.—A 2 × 2 × 2 × 4 factorial design was used to investigate the following variables: Instructions (Intentional vs. Incidental), Exposure Rate (2 vs. 8 sec.), Stimulus Pairing (Constant vs. Varied), and Number of Exposures (one, three, five, or seven).

Subjects.—An equal number (7) of introductory psychology students were randomly assigned to each of 32 treatment groups, for a total of 224 *S*s. Data for an additional 33 *S*s in the Incidental groups were omitted from the analysis because those *S*s answered "yes" when asked whether they had tried to learn the figures as well as the syllables.

Materials and apparatus.—Seven nonsense syllables were selected from Glaze's 20% association-value list. The seven geometric forms used were the same as those used and described by Fischer and Baker (1961). Forty-nine 35-mm. slides were prepared, representing each possible syllable-form combination.

The recognition test consisted of the original 7 figures interspersed among 14 other geometric figures.

Procedure.—Each *S* was run individually and was placed 15 ft. from a screen upon which the stimulus materials were projected. All *S*s were instructed to learn a serial list of nonsense syllables by the anticipation method. Intentional *S*s were further instructed to learn the figures enclosing the syllables in preparation for a recognition test of the forms. Depending upon the treatment group, *S*s were shown the slides for 2 or 8 sec. for a total of one, three, five, or seven trials. A different random order of the syllables was used for each *S*. For *S*s in the Constant-Pairing group, a given syllable appeared in a given form on all trials, whereas for *S*s in the Varied-Pairing group, a given syllable was presented in a different form on each trial. However, no form appeared more than once during any one trial.

After the requisite number of trials, all *S*s were given the recognition test and were asked to place an "x" inside each of the figures they thought they had seen during the showing of the slides. The *S*s were told that there were seven such figures, so no correction for guessing was made.

RESULTS

Figure recognition.—The mean number of forms correctly recognized by each of the treatment groups is shown in Table 1. The *S*s having only one presentation could not be divided into groups having constant and varied pairing of the stimulus materials, since the pairing was, of necessity, constant for that one trial. For this reason, the numbers of forms correctly recognized were analyzed both excluding and including the data of *S*s having had one trial. These two analyses were not construed to be independent of one another, and were done only to determine whether the effect of the relevant variable (Constant vs. Varied) was the same. The effects were almost identical, therefore only the analysis based on data of all *S*s is presented.

Each of the four variables had a significant main effect. Instructions and exposure rate were significant with $p < .001$, $F(1, 192) = 135.76$ and 12.62, respectively, while stimulus pairing and number of presentations

TABLE 1

MEAN NUMBER OF FORMS RECOGNIZED

No. of Presentations	Intentional Instructions				Incidental Instructions			
	Constant		Varied		Constant		Varied	
	2 Sec.	8 Sec.	2 Sec.	8 Sec.	2 Sec.	8 Sec.	2 Sec.	8 Sec.
1	4.14	4.71	4.29	5.29	3.29	2.70	3.43	3.14
3	5.00	6.00	4.71	5.14	3.00	3.57	3.71	4.14
5	4.14	5.57	5.14	4.71	3.57	2.43	3.43	4.14
7	4.57	5.57	5.29	6.00	3.14	4.28	3.00	4.57

were significant with a $p < .01$, F (1, 192) = 7.86, F (3, 192) = 4.89, respectively. These data indicate that instructions to learn the forms, an 8-sec. exposure interval, varying the syllable-form relationship, and increasing the number of presentations all resulted in an increased number of form recognitions. However, two of the three-way interaction effects were significant at the .05 level. The Instructions × Stimulus Pairing × Exposure Rate interaction—F (1, 192) = 4.23—is presented graphically in Fig. 1. Analysis of the simple effects of the instructions by exposure rate for constant and for varied stimulus pairing indicated the following: intentional *S*s recognized significantly more forms than did incidental *S*s, regardless of pairing ($p < .001$); with constant pairing, intentional *S*s recognized significantly more forms with an 8-sec. than with a 2-sec. exposure ($p < .025$), while incidental *S*s showed no improvement; and with varied pairing, both intentional and incidental *S*s showed a significant increase in number of forms recognized with the 8-sec. exposure ($p < .025$).

In Fig. 1, it may be noted that for the lowest level of learning (i.e., with incidental instructions and a 2-sec. exposure) and for the highest level of learning (with intentional instructions and an 8-sec. exposure), there was no facilitative effect due to varying the syllable-form relationship from trial to trial. Such facilitation occurred only for the intermediate levels.

The Instructions × Number of Presentations × Exposure Rate interaction—F (3, 192) = 2.65— is shown in Fig. 2. Analysis of the Rate × Number of Exposures interaction for the intentional *S*s indicated a significant main effect due to both rate ($p < .001$) and number of exposures ($p < .05$), with no interaction effect. Thus, *S*s instructed to learn the forms recognized significantly more of them with an 8-sec. exposure than with a 2-sec. exposure, and there was a change in performance with an increased number of exposures. The

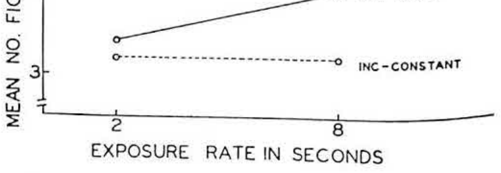


FIG. 1. Mean number of figures correctly recognized by Intentional and Incidental *S*s with Constant and Varied stimulus pairing. (Each point represents data for 28 *S*s.)

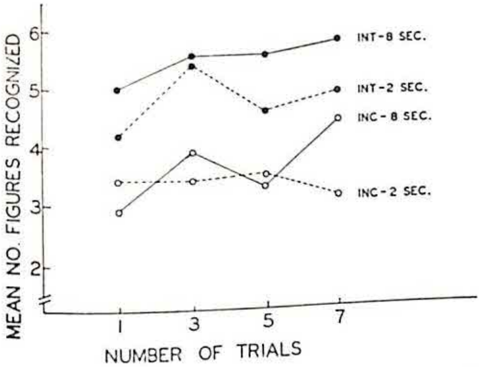


FIG. 2. Mean number of figures correctly recognized by Intentional and Incidental *S*s having a 2- and an 8-sec. exposure rate. (Each point represents data for 14 *S*s.)

Newman-Keuls method for testing differences between pairs of treatment means (using $p = .05$ as the highest acceptable value) showed that a significantly larger number of forms was recognized with three and with seven than with one trial. The other comparisons were not significant. No explanation for the large number of forms recognized by intentional *S*s having a 2-sec. exposure and three trials can be offered, except that it was a large chance difference. This explanation is supported by the nonsignificant interaction, which indicated that the deviation from parallelism of the two curves could be due to chance.

A similar analysis of the Rate × Number of Exposures interaction for incidental *S*s yielded no significant main effects but did show a highly significant difference in number of forms recognized between the 2- and 8-sec. groups who had had seven trials. In other words, the 8-sec. group finally showed a facilitative effect due to increased number of exposures, the 2-sec. group did not.

Nonsense syllable anticipation.—The mean numbers of anticipation errors per trial are shown in Fig. 3. It should be noted that the plotted points represent data for decreasing *N* as the number of trials increases.

In order to analyze the number of anticipation errors, it was decided to combine the data for all *S*s who had at least five trials. Thus the analysis is based upon the data of 14 *S*s per treatment group, or a total of 122 *S*s. Although this procedure eliminated the data of those *S*s who had only three presentations and the data on the sixth and seventh trials of those *S*s who had a total of seven presentations, it was felt that the analysis utilized the largest amount of stable information possible.

Both exposure rate and instructions had a highly significant main effect, $F\ (1, 104) = 62.90$ and 15.30, respectively, $p < .001$. Thus *S*s with an 8-sec. exposure rate learned the syllables more quickly than those with a 2-sec. rate, and *S*s given no instructions concerning the figures learned the syllables more quickly than those who were instructed to learn both figures and syllables. However, the Instructions × Exposure Rate interaction was also significant, $F\ (1, 104) = 4.61$, $p < .05$. Statistical comparison of individual pairs of means

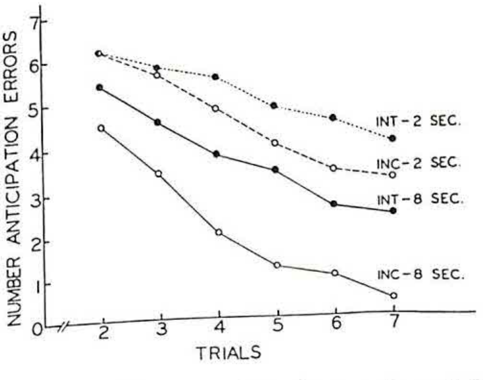


FIG. 3. Mean number of errors in anticipating nonsense syllables. (Data points for Trials 2 and 3 are based on an *N* of 21; for Trials 4 and 5 on an *N* of 14; and for Trials 6 and 7 on an *N* of 7.)

indicated that both intentional and incidental *S*s learned more syllables with an 8-sec. exposure. With a 2-sec. exposure the difference between intentional and incidental *S*s was not significant (5.62 vs. 5.19, respectively) but was highly significant with an 8-sec. exposure (4.23 vs. 2.78, respectively, $p < .001$).

The main effect of trials was also significant, $F\ (3,\ 312) = 109.97$, $p < .001$, as were the Trials × Rate, $F = 4.73$, $p < .01$, and the Trials × Instructions, $F = 8.78$, $p < .001$, interactions. The means showing the two interaction effects are presented in Table 2. Using $p = .05$ as the highest acceptable probability level, analysis of pairs of means indicated that there were consistently fewer errors with the 8-sec. exposure than with the 2-sec. rate throughout trials, and that incidental *S*s made significantly fewer errors than intentional *S*s throughout. The Newman-Keuls method showed a significant improvement for each increase in number of trials. Thus *S*s learning nonsense syllables by the anticipation method learned more with increased number of trials, but learned much more quickly with 8-sec. exposure than with 2-sec. exposure. The *S*s who had been instructed to learn figures and syllables as well as those who had no instructions concerning the figures, showed improvement with increased number of trials, but the latter *S*s learned much more quickly.

TABLE 2

MEAN NUMBER ANTICIPATION ERRORS

	Trials				
	2	3	4	5	*M*
Instructions					
Intentional	5.76	5.19	4.65	4.13	4.92
Incidental	5.32	4.67	3.38	2.60	3.98
Exposure					
2 sec.	6.24	5.80	5.14	4.48	5.40
8 sec.	4.83	4.06	2.90	2.26	3.50

TABLE 3

DISTRIBUTION OF DISCARDED INCIDENTAL *S*S

Stimulus Pairing	Exposure		
	2 Sec.	8 Sec.	Total
Constant	8	14	22
Varied	4	7	11
Total	12	21	33

Discarded Ss.—Data for 33 *S*s in the incidental groups were not included in the analysis because *S*s stated that they had tried to learn the incidental material. Distribution of the discarded *S*s with respect to exposure rate and stimulus pairing is shown in Table 3. Although the difference in cell frequencies was not significant, those *S*s with the longer exposure time and those for whom the figure-syllable relationship remained constant tended to try to learn the figures despite no instructions to do so. Since *S*s were not evenly distributed among the treatment groups, no direct analysis of number of forms correctly recognized could be made. However, in comparison to the mean number of figures recognized by the incidental *S*s retained in the analysis, 30 of the discarded *S*s recognized a number of figures greater than the mean for the corresponding treatment group. Thus the procedure of asking *S*s whether they had tried to learn the incidental material and discarding their data on the basis of an affirmative answer appears to be a necessary control in incidental-learning experiments. Omitting this procedure would tend to obscure the differences between intentional and incidental performances.

DISCUSSION

Mechanic (1962) has suggested that in a Type II design, performance of the intentional task competes with the incidental task. When the intentional and incidental materials are presented for a limited exposure interval, any experimental variable which tends to focus *S*'s attention upon the intentional stimuli would tend to reduce the proportion of exposure time available for responding to the incidental material, and result in lesser incidental learning. On this basis, it would be expected that with a longer exposure interval, a greater proportion of the time would be available for the incidental material, and greater incidental learning would occur. Such was not the case. The *S*s who had no instructions concerning the figures, did not recognize more figures with an 8-sec. than with a 2-sec. exposure interval, except when the serial order of the forms (or the form-syllable relationship) varied from trial to trial. Thus it appears that even though more time may be available for responding to the incidental material, *S* will not necessarily utilize that time unless some variable is introduced which serves to focus *S*'s attention upon the incidental material. In the present experiment, varying the order of the incidental items appears to have been an "attention-focusing" variable. In intentional learning of both the syllables and the forms, a longer exposure interval clearly facilitated learning.

According to task competition, it might also be expected that a greater proportion of the exposure interval would be available for response to the incidental material with an increased number of trials, i.e., with repeated trials, during which time *S* is learning the intentional material, *S* may spend a decreasing portion of the exposure interval on the intentional material and, consequently, more upon the incidental items. It would, therefore, be expected that incidental learning would increase with an increased number of presentations. This was found to be true only with an 8-sec. exposure and then only after seven trials. This finding may be related to the extent of intentional learning. It will be remembered that learning of the syllables occurred significantly more quickly with an 8-sec. than with a 2-sec. exposure, suggesting that *S* was spending the greater portion of the 2-sec. exposure in responding to the syllables even after as many as seven trials. Whether the task-competition interpretation is correct or not, it seems clear that the exposure rate used determines the number of trials necessary before facilitation of incidental learning occurs in a Type II design, with nonverbal incidental material. On the other hand, an increased number of trials tended to facilitate intentional learning of both verbal and nonverbal materials, with either the 2- or 8-sec. exposure.

The data of *S*s instructed to learn both the syllables and the forms appear to be consistent with the task-competition hypothesis. It seems likely that instructions to learn the forms would result in proportionately less of the exposure time being spent in responding to the syllables and more to the forms than for *S*s with no figure instructions. Thus, fewer syllables and more forms were learned by intentional *S*s, and the inferiority of syllable learning as a function of figure instruction is more pronounced with an 8-sec. exposure than with a 2-sec. exposure. Syllable anticipation was, of course, the primary task and required immediate responses. Therefore, with a 2-sec. exposure it seems likely that most of the rather limited time was spent in responding to the syllables rather than the forms, resulting in relatively less task competition. With the 8-sec. exposure, more time was available for responding to the forms, resulting in more competition with the learning of the syllables, compared to *S*s who had no figure instructions.

The effect of varying the form-syllable relationship facilitated figure recognition for intentional *S*s only with a 2-sec. exposure. It would seem that with a relatively brief exposure, the introduction of an "attentional" variable is helpful for learning, but that with the longer 8-sec. exposure, it is not necessary.

With regard to the effect of varying what might be thought of as either the form-syllable configuration or the serial order of the figures, it might be profitable to determine which variation is important, i.e., whether facilitation depends upon configurational or positional variation, or both. Other lines of investigation which might be fruitful to pursue are whether the facilitative effect of variation would occur with verbal as well as nonverbal material and whether it would occur in recall as well as in recognition.

REFERENCES

Brown, G. H. Factors influencing incidental learning. *J. exp. Psychol.*, 1954, **47**, 163–169.

Fischer, G. J., & Baker, M. R. The influence of distribution of practice and varying speeds of stimulus presentation on incidental learning. Paper read at Midwestern Psychological Association, Chicago, May 1961.

Mechanic, A. The distribution of recalled items in simultaneous intentional and incidental learning. *J. exp. Psychol.*, 1962, **63**, 593–600.

Neimark, E., & Saltzman, I. J. Intentional and incidental learning with different rates of stimulus presentation. *Amer. J. Psychol.*, 1953, **66**, 618–621.

Postman, L., & Adams, P. A. Studies in incidental learning: VII. Effects of frequency of exercise and length of list. *J. exp. Psychol.*, 1958, **56**, 86–93.

Rosenberg, S. Exposure interval in incidental learning. *Psychol. Rep.*, 1959, **5**, 675.

Saltzman, I. J., & Atkinson, R. L. Comparisons in incidental and intentional learning after different numbers of stimulus presentation. *Amer. J. Psychol.*, 1954, **67**, 521–524.

(Received May 2, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 263–267

ADAPTATION IN THE PERCEPTION OF ROTARY MOTION [1]

J. RAPOPORT

Psychological Institute, Uppsala, Sweden

The effect of adaptation to a rotary pattern on speed ratings of the pattern was studied using a stimulus known to give a clear negative aftereffect of motion when tested on the stationary condition. A standard speed was presented for every rating. When test speed was the same or less than adaptation speed and in the same direction, test speed was rated less than control speed, consistent with Gibson's hypothesis that adaptation produces a negative shift along the subjective stimulus continuum. However, when test speed was greater than adaptation speed, then the test speed was rated relatively higher. When adaptation and test speeds were in opposite directions, the results were not as clear, but some aftereffects appeared also to be negative and some positive.

The aftereffect due to viewing a rotary stimulus has usually been studied with a stationary test pattern. If one first views a rotating pattern, a stationary pattern will then appear to turn in the opposite direction. The best known example is the spiral aftereffect, but other patterns have been shown to give clear and less ambiguous aftereffects (Wohlgemuth, 1911). Johansson (1956) measured aftereffect when the circular test pattern was moving slowly in the same direction and the velocity was adjusted by *S* so that it appeared equal to the aftereffect velocity. Gibson (1937) demonstrated aftereffect on a test pattern moving at the same velocity as the adaptation pattern and obtained a phenomenal decrement in velocity, which he considered part of a shift in the correspondence between perceived velocity and all of the values along that stimulus dimension of velocity (Gibson, 1959).

Carlson (1962) tested the effect of adaptation to orbital velocity on test speeds greater, less, and the same as the adapting speed and in the same and opposite directions. He found that adaptation decreased the perceived test velocity when the objective test speed was the same or less than the adaptation speed and in the same direction. However, for test speeds greater than the adapting speed he did not obtain the negative shift predicted by Gibson's theory of adaptation. Instead, he found either an increase or no effect on test speeds higher than the adapting speed and in the same direction, and either a tendency to decrease or no effect for test speeds in the opposite direction.

Carlson used a somewhat specialized stimulus situation not commonly associated with the study of motion aftereffect. The aim of the present study is to test the generality of Carlson's findings, utilizing a rotary stimulus more similar to the usual stimulus for motion aftereffect.

METHOD

Velocity rating.—The *S* was seated in front of a viewing tunnel through which a striped

---

[1] This investigation was supported by Public Health Service Fellowship MPD 15,590 from the National Institute of Mental Health, Public Health Service.

The author wishes to thank Ingmar Dureman and the staff of the Psychological Institute for their generous help with this work.

disc could be seen which rotated at a constant speed. After 30 sec., there was a 1-sec. dark interval followed by the 2-sec. test period in which the disc reappeared surrounded by an outer ring of dots. The *S* was told that the velocity of the outer ring would be called "10" and that he must make a magnitude estimate of the velocity of the inner disc relative to the outer ring. The rating was for the 2-sec. period only, not the preceding 30-sec. period. The *S*s were instructed to fixate the center of the disc throughout the trial.

Adaptation.—For adaptation trials, the disc rotated continuously during the 30-sec. period. Control trials were the same as the adaptation trials except that the pattern was stationary for the 30-sec. period and was started during the 1-sec. dark interval. Thirty seconds elapsed between trials (whether adaptation or control).

Schedule.—The speeds of rotation were 18°, 36°, and 72°/sec (this measure refers to degrees of arc/sec for the disc rotation, not to visual angle). The standard rotated at 10°/sec and in the same direction as the stimulus disc. The experimental schedule consisted of six control conditions (each of the three speeds clockwise and counterclockwise) and 12 adaptation trials on each of 2 different days. Clockwise and counterclockwise adaptation trials were alternated. The control trials were dispersed throughout the schedule. Each *S* was presented with a different order of occurrence of conditions; on the second day each was presented his schedule of the first session in reverse order. About 2 days intervened between the two test sessions. The *S*s were divided into three experimental groups, each receiving one of the three test speeds (18°, 36°, or 72°/sec) as the adaptation speed for every adaptation trial.

Subjects.—Twenty-five male and 31 female university students served as *S*s as part of the requirement for students majoring in psychology. The data for males and females did not differ in any consistent way and were therefore combined within each of the three experimental groups.

Apparatus.—The *S* fixated binocularly through a 40-cm. viewing tunnel attached to a tachistoscope. The tachistoscope had two compartments, A and B, separated by a diagonal semitransparent mirror between opposite corners of a 45-cm. square tachistoscope box. The stimulus disc and standard ring were constructed of cardboard and were mounted on the shafts of Motors A and B, respectively. The disc stimulus consisted of a pattern of 16 equal black and white sectors. The disc diameter was 3 cm. The standard ring was a black cardboard disc with a ring of 20 white dots 3 mm. in diameter. The diameter of the ring was 6.5 cm. The shafts of the motors fitted through the walls of the tachistoscope so as to display the stimulus disc alone when Compartment A was illuminated, and the standard ring surrounding the stimulus disc when both Compartments A and B were illuminated. Some background tecture of black cardboard could be seen behind the disc stimulus when it was presented alone.

Motor A rotated at 18°, 36°, or 72°/sec. Motor B always rotated at a speed of 10°/sec. The standard ring always rotated in the same direction as the disc during the 2-sec. test period in which they appeared together. A timer controlled the lamps for the two compartments, first illuminating Compartment A for 30 sec. and, after a 1-sec. dark interval, illuminating both compartments for a period of 2 sec. Both compartments were dark between trials.

Preliminary work.—Preliminary experiments with 11 *S*s showed that the stimulus disc gave a strong negative aftereffect of motion when tested on the stationary condition (average duration 21 sec. after 30 sec. stimulation). The ratings of the test speeds by these *S*s were an approximately linear function of speed, although for speeds close to that of the standard there was a tendency for speeds to be called "10" (the speed of the standard) somewhat too frequently.

RESULTS

Aftereffect scores were calculated for each *S* by subtracting his control rating of the disc speed from his rating of that speed after the adaptation speed had been presented. A positive score indicates that he rated the speed faster after adaptation, a negative score that he rated the speed slower. Clockwise and counterclockwise rotations had no differential effect, so they have been combined. The mean of the differences in rating (adaptation minus control) was found for each group. Figure 1 shows the differential effects of the three adaptation speeds on the speed ratings of the test pattern, for speeds in the same and opposite directions from the adapta-

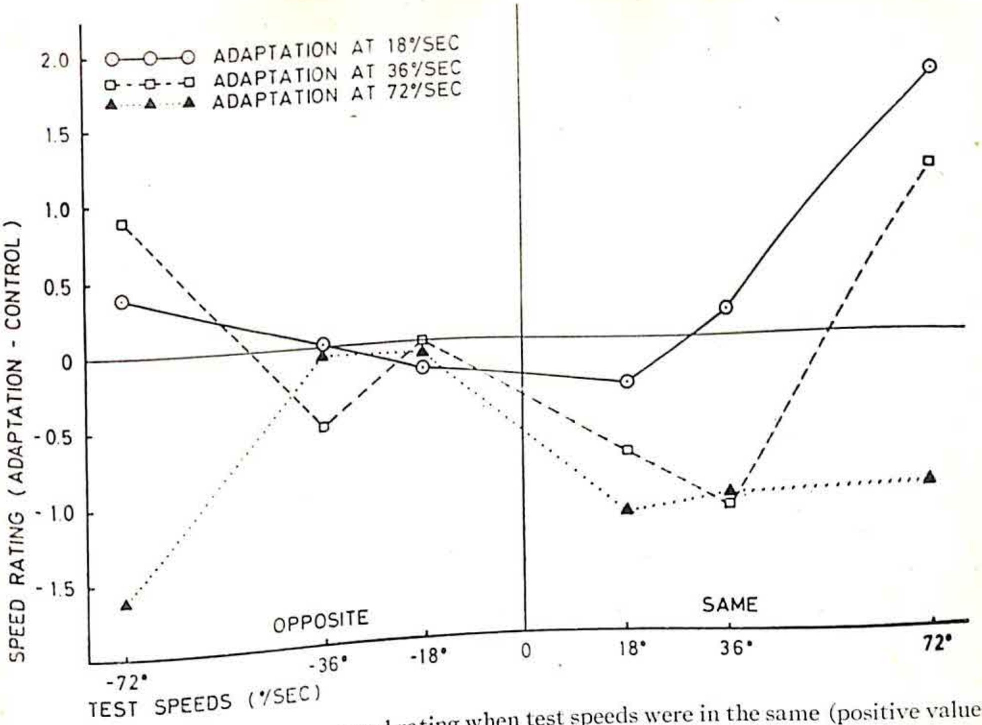


FIG. 1. Effect of adaptation on speed rating when test speeds were in the same (positive values) and opposite (negative values) direction from the adaptation speed.

tion speed. Each line represents the values for a different group of *Ss*.

The results showed a negative effect of adaptation when the test speeds were the same or less than the adaptation speed and in the same direction. This is the direction predicted by Gibson's hypothesis. However, when the test speed was greater than the adaptation speed and in the

TABLE 1

AVERAGE DIFFERENCES IN SPEED RATING (ADAPTATION MINUS CONTROL) FOLLOWING 30-SEC. ADAPTATION PERIOD

Adaptation Speed	Measure	Test Speed Same Direction			Test Speed Opposite Direction		
		18°/Sec	36°/Sec	72°/Sec	18°/Sec	36°/Sec	72°/Sec
18°/sec (*N* = 16)	*M*	−0.32	0.17	1.78	−0.18	0.01	0.40
	SD	0.82	1.09	2.88	0.86	1.55	2.44
	p	<.10	*ns*	<.02	*ns*	*ns*	*ns*
36°/sec (*N* = 20)	*M*	−0.78	−1.15	1.11	0.00	−0.54	0.90
	SD	1.24	1.42	2.20	1.21	1.56	1.63
	p	<.01	<.001	<.02	*ns*	<.10	<.05
72°/sec (*N* = 18)	*M*	−1.20	−1.08	−1.05	−0.09	−0.07	−1.62
	SD	1.24	1.05	2.26	1.28	1.61	3.43
	p	<.001	<.001	<.05	*ns*	*ns*	<.05

Note.—All probabilities derived from *t* test for difference between correlated means (single-tailed values for test speeds in same direction).

same direction, the difference in the ratings was positive (test speed rated relatively faster following adaptation). This increase in test rating was significant for two of the three conditions where the test speed was objectively faster than the adaptation speed (Table 1). When the test speed was 72°/sec and the adaptation speed was either 18° or 36°/sec, the test speed was rated significantly faster ($p < .02$). When the test speed was 36°/sec and the adaptation speed 18°/sec the difference was positive but not significant. These effects would be expected on the basis of the previous experiments cited.

When test speeds were in the opposite direction from the adaptation speed, the ratings did not differ significantly from the controls for test speeds −18° or −36°/sec. The fastest speed, −72°/sec, was rated differently from the controls by two of the three groups. For eight of the nine points with test speed in opposite direction to adaptation speed, the direction of change in rating was the same as when test and adaptation speeds were in the same direction. The average ratings for test speeds in the same and opposite directions were compared with each other (instead of with the controls), and the differences were significant for the middle and lowest test speeds but not for the highest.

The control ratings of the three groups were almost linear as a function of speed. Even though each group had a different speed for adaptation trials, the control ratings of the three groups did not differ significantly from each other ($p = .10$ for the greatest difference). There was no consistent trend in control ratings either within a session or between the first and second sessions.

DISCUSSION

Motion aftereffects to rotary stimuli when tested on the stationary condition clearly represent a negative aftereffect of motion. The present stimulus has also been shown to give a clear negative aftereffect when tested in this manner. When the test speeds were the same or less than the adaptation speed, and in the same direction, the results can also be included in the class of aftereffects which follow Gibson's hypothesis that adaptation produces a negative shift along the stimulus continuum.

The increase in apparent speed when the test speed was higher than the adaptation speed would not be predicted by Gibson's hypothesis. Carlson (1962) also found a significant increase in perceived speed, when adaptation speed was 30°/sec and test speed 53°/sec. The speeds in his study which were closer together did not produce this increase in perceived speed. In the present study adaptation at 18°/sec did not produce a significant increase in rating of the 36°/sec speed, although the difference in rating was positive. Adaptation at both 18°/sec and at 36°/sec produced an increase in rating of the test speed at 72°/sec. It would seem that adaptation and test speeds must be separated by some minimum amount before the apparent increase in speed can take place.

Adaptation-level theory might account for these results, and the findings of Helson and Nash (1960) are of particular interest in this respect. They found that an optimal difference between background and test weights was necessary before the maximal effect of background upon weight judgments was obtained. However, there is some difficulty in applying adaptation-level theory to the present study, because *S*s always made their judgments relative to a standard velocity presented on each trial. If the standard were affected by adaptation level, then the control ratings of the three groups (who presumably had three different adaptation levels) should differ; this was not the case. If the standard was not affected, then the effects of adaptation level were limited to the retinal area stimulated by the disc. It is not clear that the latter formulation would be compatible with adaptation-level theory.

For test speeds in the opposite direction the results were also at variance with any simple concept of negative aftereffect. Gibson's principle predicts an increase in apparent speed for test speeds in the opposite direction from the adaptation speed. An increase occurred only for the highest speed in the opposite direction, and then only in the groups for which an increase was also present when test and adaptation speeds were in the same direction. Nondirectional speed effects alone cannot be postulated, since the middle and lowest test speeds were rated significantly differently when test ratings for the same and opposite directions were compared. Possibly both speed and velocity factors were present which opposed each other for the middle and lowest test speeds in opposite directions.

There is some evidence for the occurrence of nondirectional speed effects. Held and White (1959) measured the perceived speed of a test line after inspection of visual noise on the face of a TV tube and found a decrease in perceived speed. After *S* was exposed to a hyperstable field (a photograph of the TV noise), they obtained an increase in perceived speed of the test line. It would be of importance for further study of velocity perception to elucidate the precise nature of such nondirectional speed effects and the conditions favoring them.

REFERENCES

Carlson, V. R. Adaptation in the perception of visual velocity. *J. exp. Psychol.*, 1962, **64**, 192–197.

Gibson, J. J. Adaptation with negative aftereffect. *Psychol. Rev.*, 1937, **44**, 222–244.

Gibson, J. J. Perception as a function of stimulation. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 1. New York: McGraw-Hill, 1959. Pp. 565–621.

Held, R., & White, B. Sensory deprivation and visual speed: An analysis. *Science*, 1959, **130**, 860–861.

Helson, H., & Nash, M. C. Anchor, contrast, and paradoxical distance effects. *J. exp. Psychol.*, 1960, **59**, 113–121.

Johansson, G. The velocity of the motion after-effect. *Acta psychol., Amsterdam*, 1956, **12**, 19–24.

Wohlgemuth, A. On the aftereffect of seen movement. *Brit. J. Psychol. monogr. Suppl.*, 1911, **1**, 1–117.

(Received May 8, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 268–275

REACTION TIME UNDER THREE VIEWING CONDITIONS: BINOCULAR, DOMINANT EYE, AND NONDOMINANT EYE[1]

PATRICIA KELSEY MINUCCI AND MARY M. CONNORS

United States Naval Medical Research Laboratory, Groton, Connecticut

Reaction times for 10 Os were measured at 4 photopic intensity levels: 7.13, 8.06, 9.10, and 10.18 log $\mu\mu$L., under binocular, dominant eye, and nondominant eye viewing conditions. Reaction time has been found to be a negatively accelerated, decreasing function of increased light intensity. There is a constant relationship between the binocular reaction time and the average monocular reaction time for each O. Equivalent brightnesses under each viewing condition were calculated on the basis of speed of reaction. The resulting curves suggest that binocular reaction times are faster than would be expected from the data of the dominant and the nondominant eye, even assuming complete summation.

A continuing problem in visual psychometrics is concerned with how to measure changes in effect when the stimulus is varied over a range of intensities. Reaction time (RT) has shown promise as a sensitive measure at levels above threshold. Studies involving change in reaction time with change in intensity of the visual stimulus (Bartlett & Macleod, 1954; Froeberg, 1907; Hovland, 1936; Steinman, 1944; Woodrow, 1915) agree that reaction time decreases with an increase in intensity, but they disagree on the amount of the change. They agree that there is a region of intensity in which the curve flattens, but they differ in their determinations of the region. In many of the early studies the experimental variables are not stated, the number of *O*s was small, and when several intensities were used, often two or more log steps were omitted. These factors undoubtedly contributed to the disagreements. However, since reaction time can be shown to vary inversely with the intensity of visual stimulus, a new approach is opened to the problem of the relative effectiveness of monocular and binocular stimulation.

Several studies, using various techniques, (Bartlett & Gagné, 1939; Collier & Kubzansky, 1958; Crawford, 1940; Crozier & Holway, 1939; Forbes & Mote, 1956; Graham, 1930; Lythgoe & Phillips, 1938; Pirenne, 1943) have investigated summation at threshold. Their findings ranged from no summation to complete summation.

Experiments designed to measure binocular summation at levels above the threshold have shown more general agreement. Brightness matching, critical flicker frequency (cff), and reaction time have been used to measure the effect. Sherrington (1904-05), Graham (1930), and Fry and Bartley (1933) found that suprathreshold stimuli viewed binocularly appeared brighter than when viewed monocularly, although there is considerable difference in the amount of summation. Leibowitz and Walker (1956) studied the effect of field size and luminance on the binocular summation of stimuli and found that the test field appeared brighter when viewed binocularly than when viewed with one eye. They also found an increase in summation with increase in stimulus size. When he measured cff, Ireland (1950) found that in-phase binocular stimulation yielded higher values and out-of-phase stimulation lower values than those obtained with monocular stimulation. Since Poffenberger's study of 1912, it has been generally accepted that binocular reaction time is faster than monocular to the same visual stimulus, although the reliability of his results have been questioned (Teichner, 1954, p. 131). In a study involving discrimination of a ring from its surround, Dawson (1913) found that as the intensity of the ring and the background decreases, the RTs increase and the binocular RT was much lower than the monocular. Travis (1949) measuring the discrimination for the break in a Landolt ring, found that binocular RT was lower than monocular. Smith (1955) found binocular RT 15–19% faster than monocular depending on the luminance level. He found further that raising the intensity 1 log unit resulted in a decrease in RT of approximately 25% for both monocular and binocular vision.

---

[1] This paper was adapted from the thesis of Patricia Kelsey Minucci, submitted in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Psychology at the University of Connecticut in 1959.

The opinions or assertions contained herein are the private ones of the authors and are not to be construed as official or reflecting the views of the Navy Department or the Naval Service at large.

In the present study, we propose first to clarify the relation of reaction time to stimulus intensity over a range of 3 log units, and then to answer the question of whether there is summation between the eyes beyond that which would be expected from chance alone. It will be a further purpose of this study to determine if there is any significant difference between the dominant and the nondominant eye as measured by reaction time.

METHOD

Apparatus.—The *O* was positioned with his head at the center of a hemispherical screen, facing into the screen. His head was steadied by an adjustable chinrest. A white metal occluder, pivoted by one edge, could be rotated to cover either eye or to leave both eyes uncovered. An *O* was carefully aligned at the beginning of a session to make sure that the stimulus, lying at the apex of the hemisphere, was oriented symmetrically to both eyes. The white interior of the hemisphere was lighted to 6.1 log $\mu\mu$L. Translucent diffusing glass covered a hole of 1°

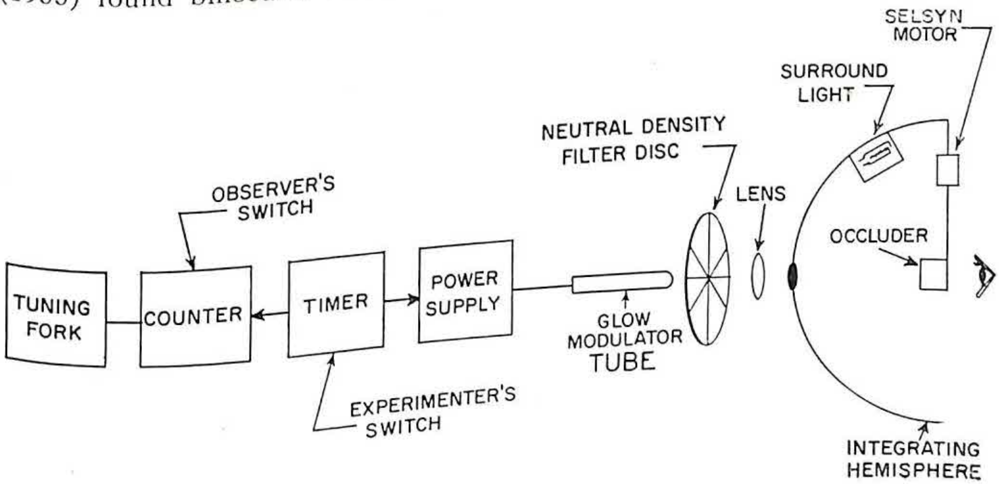


FIG. 1. Diagram of the apparatus.

diameter in the apex of the hemisphere. The stimulus was projected on the rear of this diffusing glass. Between stimulations the area appeared somewhat darker than the screen and was clearly distinguishable from it.

The stimulus light source was a Sylvania glow-modulator tube, R1131C, regulated to give a flash of 60 msec. A biconvex lens focused the light from the tube onto the rear of the diffusing glass. A series of neutral density filters, mounted in a disc, permitted presentation of the various stimulus intensities. A diagram of the apparatus is given in Fig. 1. The light source was connected in series with a Hunter timer and a four-sequence predetermined Potter counter with an input signal from a 1,000-cps tuning fork. In the circuit were two switches, one for *O* and one for *E*. When a verbal signal was given, *O* depressed his switch and held it until he saw the stimulus flash. After a 1–2 sec. delay *E* engaged his switch, the mechanism was activated, i.e., the timer engaged the glow-modulator tube, and the counter was started. The *O* stopped the counter by releasing his switch and thus recorded his reaction time.

Procedure.—Ten *O*s, corrected when necessary to 20/20 vision, served in this experiment. Each *O* was given three conventional tests of eye dominance. These tests all involved sighting tasks. For the purpose of this experiment, "eye dominance" is defined as simply the results obtained from these tests. The results were sufficiently consistent to allow ready classification of each *O* as "left-" or "right-" eye "dominant."

A session consisted of one or two predetermined series involving three presentations at each of four light levels: 10.18, 9.10, 8.06, and 7.13 log $\mu\mu$L. under each viewing condition: dominant eye, nondominant eye, and both eyes. In each series both the light levels and the viewing conditions were randomized so that except for an occasional replication, each successive stimulus was different from the preceding one. As many practice sessions were run as necessary before data were taken. The mean of 10 reaction times for each condition was used in the analysis of variance. Since trained *O*s were employed, no catch trials were included in the actual experiment and none of the *O*s made any anticipatory responses.

RESULTS

The mean reaction times for each *O* at each experimental condition are given in Table 1. The RTs, averaged for all *O*s, as a function of intensity, are shown in Fig. 2A. The data were then recast in terms of the averaged reciprocal of RT, according to the method suggested by Piéron (1952, p. 354) and are shown in Fig. 2B.

TABLE 1

MEAN REACTION TIME IN MSEC. FOR EACH *O* UNDER EACH STIMULUS CONDITION

O	Dom. Eye	Stimulus Intensities in Log $\mu\mu$L.			
		7.13	8.06	9.10	10.18
		Dominant Eye			
MC	R	280.0	251.3	220.1	208.8
BH	L	270.2	244.0	232.9	224.9
JK	L	268.5	256.0	217.9	208.4
PK	R	313.6	291.2	260.8	243.2
SL	R	269.8	228.4	211.6	197.8
GM	R	310.8	273.8	251.8	233.2
AR	L	482.7	371.7	338.9	331.8
HS	R	294.7	274.3	245.3	229.0
IS	L	227.2	205.1	192.1	181.5
JV	R	362.6	293.0	260.9	247.2
M		308.01	268.88	243.23	230.58
		Nondominant Eye			
MC		306.6	255.2	229.3	224.8
BH		304.7	247.5	230.1	226.7
JK		308.5	265.3	234.7	212.6
PK		353.0	301.9	274.4	254.3
SL		290.6	232.3	212.2	204.6
GM		317.0	278.0	254.3	236.9
AR		484.6	392.7	354.6	349.6
HS		300.1	290.8	250.6	227.0
IS		235.5	208.6	196.7	183.8
JV		399.6	317.1	282.2	269.6
M		330.02	278.94	251.91	238.99
		Both Eyes			
MC		249.2	233.8	218.2	200.6
BH		270.3	238.5	222.2	209.8
JK		253.3	231.1	218.8	203.8
PK		286.2	266.1	260.1	238.3
SL		259.4	229.2	207.8	195.5
GM		294.0	265.3	234.2	229.1
AR		426.3	357.3	336.0	322.7
HS		265.1	241.3	232.5	217.3
IS		211.5	188.8	180.8	171.6
JV		297.8	260.3	238.8	232.9
M		281.31	251.17	234.94	222.16

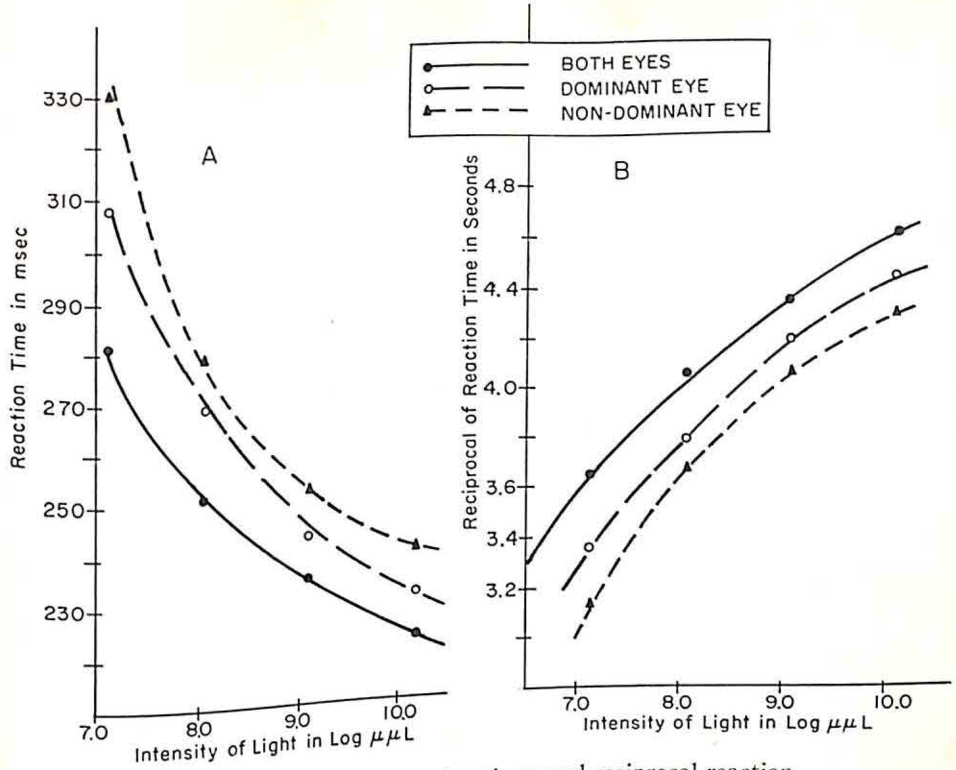


FIG. 2. Averaged reaction times and reciprocal reaction times as a function of intensity.

The data from the present study show that RT is a negatively accelerated, decreasing function of increasing log intensity. Or, phrased another way, the reciprocal of RT increases with the increase in log intensity. This conclusion is valid over the range of photopic stimuli presented in this study and for each of the viewing conditions.

Binocular RTs are faster than dominant eye RTs which in turn are consistently faster than the non-dominant eye RTs at all intensity levels. The overall increase in RT from binocular to dominant eye is approximately 6%, and to the non-dominant eye is 10%. There is a constant relationship between the amount of the binocular RT and the average monocular RT for each *O*. The rank-order correlation is .95, which is highly significant. This means that, knowing the regression equation, one could accurately predict the average monocular RT from the binocular RT for any individual.

TABLE 2

ANALYSIS OF VARIANCE OF RECIPROCAL REACTION TIMES FOR 10 *O*s AT FOUR INTENSITY LEVELS AND WITH THREE VIEWING CONDITIONS

Source of Variation	*df*	*F*
*O*s	9	253.049**
Intensities (I)	3	49.394**
Viewing conditions (V)	2	3.529**
*O*s × I	27	4.100**
*O*s × V	18	4.614**
I × V	6	
*O*s × I × V	54	
Within (*MS*)	1,080	(0.150)
Total	1,199	

** $p < .01$.

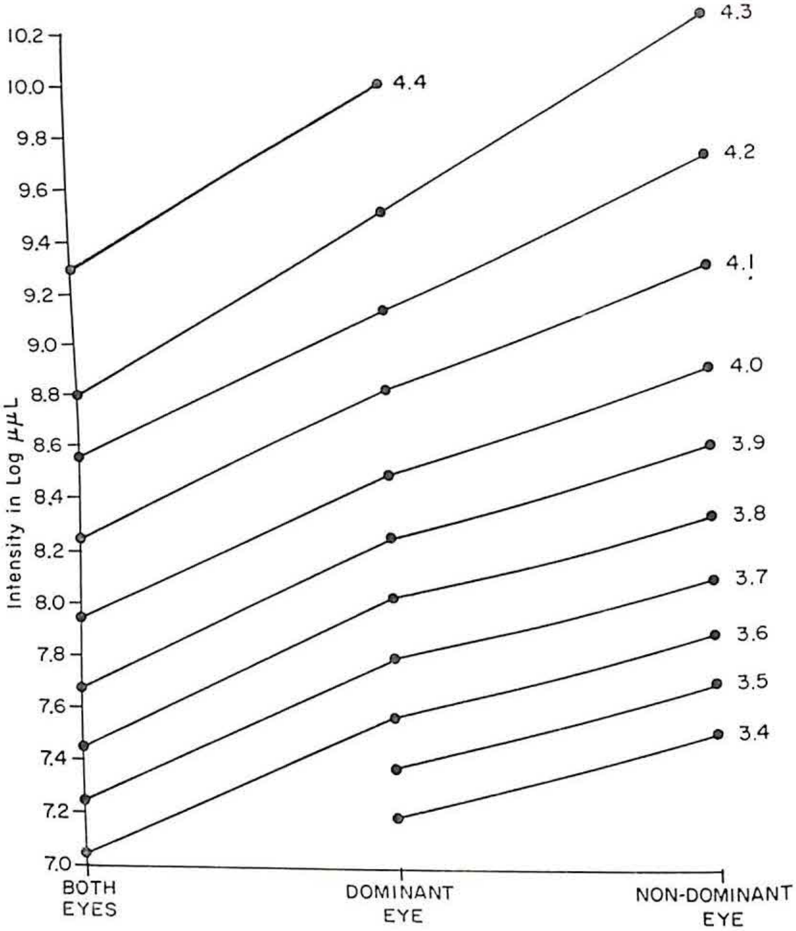


FIG. 3. Brightness required to produce equal reaction times under the three viewing conditions.

Table 2 gives the results of the triple-classification analysis of variance. All main effects and double interactions are significant at $p < .01$. Although there is no proper test of the individual differences, the two double-interaction terms involving this variable can be pooled to test the *O* mean square, and the resulting *F* of 88.5 is, of course, highly significant. Interpretation of the nonsignificant triple interaction is that the interaction between two variables (for example, *O*s and Intensities) is not varying according to the values of the third variable (monocular or binocular viewing condition).

Figure 3 shows the brightnesses required to produce the same RTs under the three viewing conditions. Each curve was obtained by reading the log brightness at each viewing condition across each reciprocal RT from the curves of Fig. 2. The resultant curves in Fig. 3 may be thought of as equal-brightness-effectiveness contours. These contours show that, for a visual stimulus to evoke the same speed of reaction (equal latency), the light must be between .53 and .71 log unit higher on the brightness scale when presented to the dominant eye than it is when presented to both eyes. At the lower intensities, the same speed of reaction is evoked when the light is .85 log unit brighter, presented to the nondominant eye, and at high intensities when the light is 1.4 log units brighter, presented to the nondominant eye, than to both eyes. At the high brightnesses, the stimulus had to be .72 log unit brighter to evoke the same response in the nondominant as in the dominant eye, while at the intermediate levels, the difference in equivalent brightness averaged .35 log unit.

In order to obtain an index of the strength of the dominant eye in each *O*, the dominant vs. nondominant differences, averaged for an *O* over all intensities, were ranked according to the percent decrease in reciprocal RT. This index was subsequently correlated with several other measures taken prior to the actual experiment: binocular visual acuity, visual acuity differences between the dominant and the nondominant eyes; as well as with binocular RT, and the age of *O*. None of these correlations proved to be statistically significant.

DISCUSSION

The data of the present experiment agree with those of previous studies that RT varies inversely with stimulus intensity. Although a comparison of the absolute amounts of RT found in the various researches is difficult due to the different, and sometimes unknown, conditions of previous works, some comparisons can be made.

There is a 25% decrease in RT from the overall average at 7.13 log $\mu\mu$L. to the average at 10.18 log $\mu\mu$L., which is comparable to that found by Woodrow (1915); while Steinman (1944) found an 11.74% decrease from 6 to 9 log $\mu\mu$L. Bartlett and Macleod (1954) concluded that the curve was relatively flat between 8 and 12 log $\mu\mu$L. The present study shows that RT continues to decrease significantly from 8 to 10 log $\mu\mu$L. with no indication that it would not continue to decrease beyond the 10.18-log-$\mu\mu$L. level. This is the area where Steinman (1944) found an abrupt increase in RT for two *O*s.

At the intensity levels used, summation is clearly demonstrated. The percent decrease in binocular RTs as compared with the average monocular RTs ranges between 5% at the highest intensity and 12% at the lowest intensity. Using a 2-cp. light, Poffenberger (1912) found an 8% decrease in binocular RTs vs. monocular RTs. As shown in Table 3 a light stimulus to both eyes had the same effect as one ranging between .69 and 1.07 log units brighter to one eye, depending on the luminance level. It should be noted that the

TABLE 3

MONOCULAR-BINOCULAR AND NONDOMINANT-DOMINANT DIFFERENCES IN LOG $\mu\mu$L. FOR THE RANGE OF RECIPROCAL REACTION TIMES

Reciprocal of Reaction Time	Differences Av Mono-Binoc	Ratio Av M/B	Differences N-Dom-Dom	Ratio N-D/D
3.4			.31	1.043
3.5			.31	1.042
3.6	.69	1.098	.31	1.041
3.7	.70	1.097	.30	1.038
3.8	.74	1.099	.30	1.037
3.9	.76	1.099	.33	1.040
4.0	.76	1.096	.40	1.047
4.1	.82	1.099	.45	1.051
4.2	.87	1.101	.55	1.060
4.3	1.07	1.121	.72	1.076

absolute log-unit difference varies according to the intensity level. However, as previously stated with reference to RT, there is a constant relationship between average monocular and binocular brightness equivalents. For the intensities of this study, the constant approximates 1.1. The log-unit differences for equivalent RTs between the average monocular and binocular conditions are greater than could be expected even with complete summation. This indicates that the increase in brightness with binocular vision only partially explains the reductions in reaction time.

Unlike the monocular-to-binocular ratios, the nondominant-to-dominant ratios of Table 3 are not constant and indicate that the differences between the nondominant and the dominant eye as measured by reaction time are slightly reduced in the middle range of intensities. At any intensity level, these differences are considerable and imply that when only one eye is being used for observations, the factor of dominance is an important one.

Contrary to what might be expected, dominant eye visual acuity was superior to nondominant eye acuity in only four cases, the same in one, and in five cases acuity was superior in the nondominant eye. It is also interesting to note that binocular visual acuity was superior to that of the dominant eye in only six *O*s. It is concluded that acuity is not a factor in determining the superiority of the binocular RT over monocular, or of the dominant eye RT over nondominant.

The *O*s in this study are all right-handed. Clinicians have felt that individuals with crossed dominance such as found in the four *O*s with left-eye dominance may be hampered in terms of the equilibrium of the most highly integrated functions. However, the RT means for these "left-eyed" *O*s were lower than the means for the "right-eyed" *O*s, although the differences were not statistically significant. This sensitive measure, speed of response, has apparently not been hampered by the factor of crossed dominance.

REFERENCES

BARTLETT, N. R., & GAGNÉ, R. M. On binocular summation at threshold. *J. exp. Psychol.*, 1939, **25,** 91–99.

BARTLETT, N. R., & MACLEOD, S. Effect of flash and field luminance upon human reaction time. *J. Opt. Soc. Amer.*, 1954, **44,** 306–311.

COLLIER, G., & KUBZANSKY, P. The magnitude of binocular summation as a function of the method of stimulus presentation. *J. exp. Psychol.*, 1958, **56,** 355–361.

CRAWFORD, B. H. Ocular interaction in its relation to measurements of brightness threshold. *Proc. Roy. Soc. Lond., Ser. B*, 1940, **128,** 552–559.

CROZIER, W. J., & HOLWAY, A. H. Theory and measurement of visual mechanisms; III. ΔI as a function of area, intensity, and wave-length, for monocular and binocular stimulation. *J. gen. Physiol.*, 1939, **23,** 101–141.

DAWSON, S. Binocular and uniocular discrimination of brightness. *Brit. J. Psychol.*, 1913, **6,** 78–108.

FORBES, L. M., & MOTE, F. A. A comparison of the variability of binocular and monocular threshold measurements during dark adaptation in the human eye. *J. comp. physiol. Psychol.*, 1956, **49,** 431–436.

FROEBERG, S. The relation between the magnitude of stimulus and the time of reaction. *Arch. Psychol., N. Y.*, 1907, **16,** 1–38.

FRY, G. A., & BARTLEY, S. H. The brilliance of an object seen binocularly. *Amer. J. Ophthal.*, 1933, **16,** 687–693.

GRAHAM, C. H. An investigation of binocular summation. I. The fovea. *J. gen. Psychol.*, 1930, **3,** 494–510.

HOVLAND, C. I. The influence of adaptation illumination upon visual reaction time. *J. gen. Psychol.*, 1936, **14,** 346–359.

IRELAND, F. H. Comparison of critical flicker frequencies under conditions of monocular and binocular stimulation. *J. exp. Psychol.*, 1950, **40,** 282–286.

LEIBOWITZ, H., & WALKER, L. Effect of field size and luminance on the binocular summation of suprathreshold stimuli. *J. Opt. Soc. Amer.*, 1956, **46,** 171–172.

LYTHGOE, R. J., & PHILLIPS, L. R. Binocular summation during dark adaptation. *J. Physiol.*, 1938, **91,** 427–436.

PIÉRON, H. *The sensations: Their functions, processes, and mechanisms.* (Trans. by M. H. Pirenne & B. C. Abbott) New Haven: Yale Univer. Press, 1952.

PIRENNE, M. H. Binocular and uniocular thresholds of vision. *Nature*, 1943, **152**, 698–699.

POFFENBERGER, A. T. Reaction time to retinal stimulation, with special reference to the time lost in conduction through nerve centers. *Arch. Psychol., N. Y.*, 1912, **23**, 1–73.

SHERRINGTON, C. S. On binocular flicker and the correlation of activity of "corresponding retinal points." *Brit. J. Psychol.*, 1904–05, **1**, 26–60.

SMITH, W. M. Effect of monocular and binocular vision, brightness, and apparent size on the sensitivity to apparent movement in depth. *J. exp. Psychol.*, 1955, **49**, 357–362.

STEINMAN, A. R. Reaction time to change compared with other psychophysical methods. *Arch. Psychol., N. Y.*, 1944, **292**, 34–60.

TEICHNER, W. H. Recent studies of simple reaction time. *Psychol. Bull.*, 1954, **51**, 128–149.

TRAVIS, R. C. Studies of visual discrimination time: The time required to recognize simple patterns at equal distances from the eye, and patterns at alternately far and near distances. *USAF AMC Aeromed. lab. Rep.*, 1949, No. 2 (TSEAA-694-1H).

WOODROW, H. Reactions to the cessation of stimuli and their nervous mechanism. *Psychol. Rev.*, 1915, **22**, 423–452.

(Received May 9, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 276–285

RELATIONSHIP BETWEEN WORD FREQUENCY AND RECOGNITION: PERCEPTUAL PROCESS OR RESPONSE BIAS?[1]

ROBERT B. ZAJONC
University of Michigan

AND

B. NIEUWENHUYSE
University of Amsterdam

Functions relating word frequency to recognition threshold and to pseudo-recognition threshold were compared for 24 Dutch Ss. Luminance thresholds were obtained to 10 Turkish words previously shown with frequencies of 1, 2, 5, 10, and 25. Pseudo-recognition thresholds were obtained in the absence of stimuli for the same words and the same Ss. Drive was manipulated to determine whether it interacts with frequency, taking this interaction as an indication of response bias. The comparison of the frequency-threshold functions, the absence of a drive-frequency interaction, and the decreasing dependence of response emission on prior training with increasing stimulus information, demonstrated that response bias plays a negligible role in the frequency-recognition relationship when stimuli are present and when a stringent recognition criterion is employed.

The relationship often found between word frequency and recognition threshold has recently been challenged by Goldiamond and Hawkins (1958) and Havens and Foote (1963). Goldiamond and Hawkins found that "pseudo-recognition" threshold, that is, one estimated from data collected in the absence of stimuli, is, like threshold obtained with conventional procedures, a negative logarithmic function of the frequency of prior experience with the stimulus words. They argue that "if the previously reported frequency-recognition relationship is now [in the absence of stimuli] obtained, this relationship cannot be ascribed to perception [p. 458]." The finding was offered as "challenging a perceptual interpretation of the word-frequency-recognition relationship, where similar procedures are utilized [p. 462]." It is thus suggested that the relationship previously reported is an artifact of response bias and essentially independent of the stimulus. Words which occur frequently in the language are more often used as guesses than infrequent words. There is, therefore, a greater likelihood for a match between *S*'s guess and the tachistoscopically exposed stimulus for frequent than for infrequent words. Since thresholds are commonly measured by the ascending method of limits, and since these high-frequency word responses are also likely to occur as guesses early in the ascending series, low thresholds to high-frequency words are spuriously obtained.

But on the basis of their data alone Goldiamond and Hawkins' conclusions are not entirely justified. Given, as they seem to propose, that response bias fully accounts for the relationship, then the frequency-recognition and the frequency-pseudo-recognition functions must be *identical*. Goldiamond and Hawkins estimated a relationship between frequency and threshold from data collected in the absence of stimuli and found it to be *similar* to that obtained in their presence, namely negative and logarithmic. But they have not determined whether it was, in fact, identical.

---

[1] This research was carried out when the major author was a Fulbright scholar at the University of Amsterdam. He wishes to express gratitude to H. C. J. Duijker for kindly making available the facilities of the Psychological Laboratory of the University of Amsterdam. The research was in part also sponsored by the Air Force Office of Scientific Research, Grant AFOSR-21-63.

That the frequency-recognition and the frequency-pseudo-recognition functions are not identical is very likely. Take a word which occurs only once in a million. If the probability of emission of this word as a guess is also one in a million,[2] then the expected mean number of trials up to and including the first correct guess would be 1,000,000 (Feller, 1950, p. 174). If we begin the ascending series with a 10-msec. exposure and increase exposures in steps of 10 msec., then the expected mean threshold would be 2 hr. 40 min.! On the other hand, a word which, for the purpose of argument, has the probability of $\frac{1}{2}$ would show a threshold of 20 msec.—too low in general for recognition to occur. Since pseudo-recognition threshold overestimates actual recognition threshold for low-frequency words and underestimates it for high-frequency words, the frequency-threshold functions obtained in the presence and in the absence of stimuli will intersect at some point between low and high word frequencies.

The present study involves a comparison of these two functions in an attempt to examine the contribution of stimulus and response factors to the relationship between word frequency and recognition threshold. Recognition and pseudo-recognition thresholds are obtained for the same set of words, differing in frequency, and for the same *S*s. Also, a threshold measure obtained in the presence of stimuli and requiring only one correct identification is compared with a measure requiring three successive correct identifications.

In a further independent and more indirect attempt to examine the role of response bias in the frequency-recognition relationship the *S*'s motivation was manipulated. There is impressive evidence consistent with the S-R drive theory (Spence, 1956) which indicates that when response tendencies are in competition an increase in drive facilitates responses governed by strong habits and impairs responses governed by weak ones (Spence, Farber, & McFann, 1956; Spence, Taylor, & Ketchel, 1956). Viewed as an artifact of response bias, the recognition situation involves competition of response tendencies varying in strength. When drive is increased the probability of high-frequency responses should increase and the probability of low-frequency responses should decrease. Because of the alleged dependence of threshold on response probability the inverse effect should obtain for threshold. Such an interaction between motivation and frequency was recently obtained for recognition data by Glucksberg (1962). Thus, if response bias *is* a significant factor in the relationship between frequency and threshold, we will find that drive interacts with frequency, and if it is not a significant factor no such interaction will obtain.

METHOD

Subjects.—The *S*s were 24 Dutch male university students recruited from the Student Baby-Sitter Service in Amsterdam. They were paid 5.00 Dutch guilders for participation in the experiment.

Stimulus words.—The stimuli were 24 Turkish words taken from a study by Solomon and Postman (1952). As in their experiment, 10 of these were *core words*, divided into five frequency classes (1, 2, 5, 10, and 25) of two words each. To counterbalance possible differences in intelligibility each core word was used in every frequency class at least twice for different *S*s. Fourteen stimuli were *filler words*, each presented once during training.

---

[2] This assumption is supported by Goldiamond and Hawkins' (1958) data. The function relating frequency of response emission to the frequency of training had a slope approaching 1.00 and an origin near zero.

Training procedure.—The core and filler words were typed in upper-case letters in the center of 4 × 6 in. white cards. For each *S* a deck of 100 cards was made up, consisting of 50 cards for the two 25-frequency words, 20 cards for the two 10-frequency words, etc., and one card each for the 14 filler words. The training instructions and procedure were similar to those of Solomon and Postman (1952), emphasizing that the experiment dealt with the learning to pronounce and read foreign words. The cards, taken one by one from the previously shuffled deck, were shown to *S* with *E* reading the word upon its presentation. The *S* was asked to look carefully at the word and to pronounce it after *E*.

After training *S* was asked to read a passage from *Taal en Psychologisch Werkelijkheid* (Duijker, 1946). This interpolated activity took approximately 5 min.

Recognition and pseudo-recognition procedure.—The 10 core words were reproduced on 2 × 2 in. transparent slides. They were projected on a 3 × 3 ft. milk-glass screen by a 100-w. projector, equipped with a timing shutter and a device for controlling luminance level. A mounting enclosing two overlapping Polaroid discs was affixed directly in front of the lens. One of the discs was stationary, while the other could be rotated in steps of .5° over an angular-displacement range of 0° to 90°. Luminance levels in foot-lamberts corresponding to the angular settings were found by measuring luminance at *S*'s side of the screen in steps of 5°, and by interpolating other values from the relationship obtained. The milk-glass screen was located 6 ft. in front of the projector and *S* sat 3 ft. behind it in an almost completely dark room. The angle subtended by the projected words was 18°.

Prior to the assessment of thresholds *S*s were given four practice words (*AMSTERDAM, LERAAR, UNIVERSITEIT, PROFESOOR*) in increasing luminance steps up to full recognition. Following practice, for each core-word luminance threshold as well as pseudo-recognition threshold were determined. For this purpose the series was divided into four trial blocks, each consisting of 100 presentations. On each trial block the 10 core words were each shown five successive times in increasing luminance steps of −.10 log ft-L and for $\frac{1}{25}$ sec. The ascending series began with an exposure of −1.82 log ft-L. Among the luminance-threshold trials were randomly interspersed an equal number of pseudo-recognition trials, on which a blank slide was projected at $\frac{1}{25}$ sec. exposure and always two steps above the minimal luminance level. Before each presentation a verbal warning was given.

For each *S* a score sheet was constructed, consisting of four parts, one for each trial block. Each part contained 20 columns and five rows. Each core word headed 2 of the columns. For 1 of the 2 columns the trials constituted ascending luminance levels of the core word, and for the other, repeated constant luminance presentations of the blank slide. The positions of the words were randomized among the columns, independently for each trial block and for each *S*. The five rows in each column constituted increasing luminance levels for stimuli actually shown and repeated trials for blanks. The session began by *E* showing the slide called for by the first column five times, either in increasing luminance steps for stimulus slides, or at the same low luminance for blanks. After the five presentations, luminance level was reset to the minimal level, and the slide called for by the second column was shown five times in succession, and so on. This procedure was repeated in each trial block, except that in the case of stimulus slides, the presentations began with a luminance level just above the last exposure of the preceding trial block. The *S*s were instructed to report the words shown and to guess when nothing or very little could be seen. All responses were recorded literally. No feedback as to whether the response was correct, or any other form of reinforcement, was given.

Three measures were obtained: the number of trials up to and including the first response which was congruent with the score sheet for blanks (SA measure); the number of trials of increasing luminance steps up to and including the first response congruent with the word actually presented (SP_1); and the number of trials up to and including the first response congruent with the word presented when that response was immediately followed by at least two consecutive congruent responses (SP_3). A response was considered congruent when it had at least five letters and at least one entire syllable in common with the core word. For the SA and SP_1 measures, when no congruent responses occurred within 20 presentations, *S* was arbitrarily assigned a score of 21. For the SP_3 measure an arbitrary score of 21 was assigned when either no congruent responses occurred, or when the congruent response occurred on the nineteenth or the twentieth trial.

Manipulation of motivation.—Twelve *S*s served in the low- and 12 in the high-drive group. Motivational level was manipulated following the training session. To the instructions of the high-drive group was added the emphasis that it was important to perform well. It was further announced that the *S* with the lowest average overall threshold will receive a 30.00-guilder prize, and that the top 25% of the remaining *S*s will each receive a 5.00-guilder prize. The prizes offered as incentive were awarded after the completion of the entire experiment.

RESULTS

Comparison of the frequency-threshold functions in the presence and absence of stimuli.—Figure 1 shows recognition thresholds for the SP_1 and the SP_3 measures and the pseudo-recognition thresholds for the SA measure. It is quite clear that all measures vary as a function of frequency. The effect of frequency is significant at the .001 level, F (4, 92) = 56.67. Taken separately, all measures have significant slopes. The effect of frequency on the SA, SP_1, and

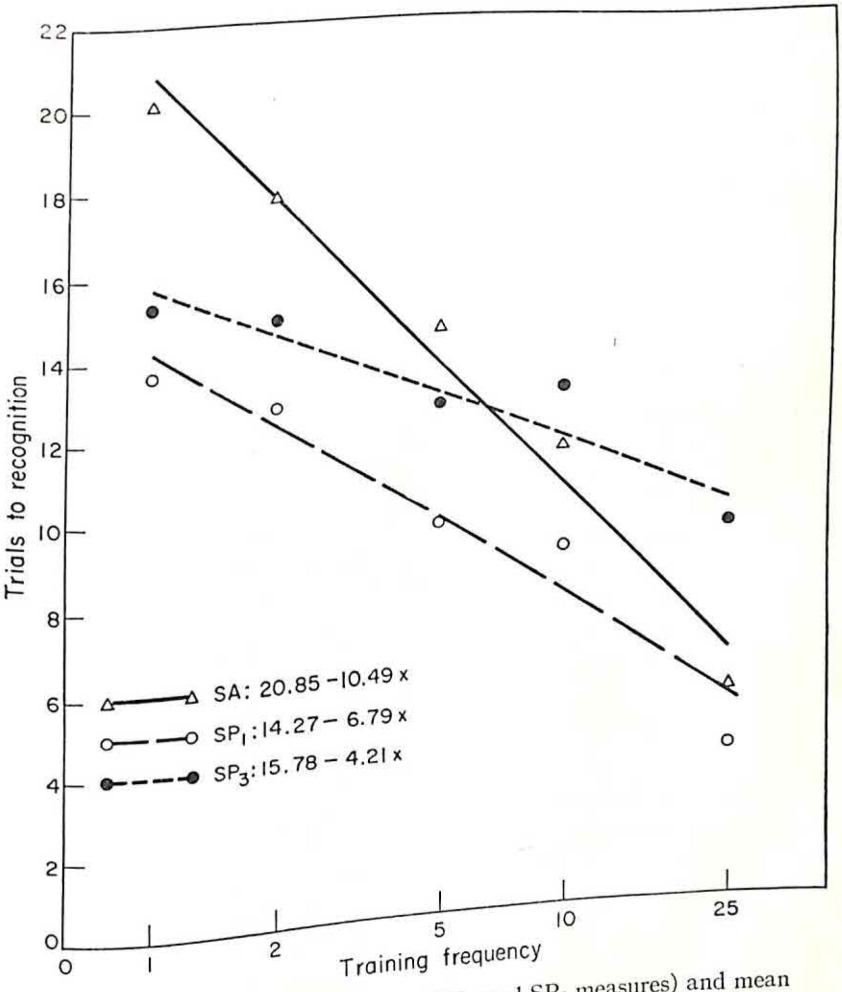


FIG. 1. Mean trials to recognition (SP_1 and SP_3 measures) and mean trials to pseudo recognition (SA measure).

SP_3 measures showed *F* ratios of 47.20, 35.16, and 19.28, respectively, all significant at the .001 level (*df* = 4/88). It is also clear that the functions of the SP_3 and the SA measures intersect as expected. The F × M interaction (Frequency × Measures) reached an acceptable level of significance with *F* (8, 184) = 11.73. The results further show that the overall threshold obtained by means of the SP_1 measure is significantly lower than that measured by either the SP_3 or the SA measure, *F* (2, 46) = 68.56.

We note from Fig. 1 that the SA and the SP_1 functions (obtained by the least-squares method) converge at high frequencies and that the SP_1 and SP_3 functions converge at low frequencies. Let us consider the high frequencies first. High-frequency responses occur often and early in the recognition series. Consequently, when only one correct identification is used as the threshold criterion (SP_1) the probability of a spurious recognition is rather high. The convergence of the SP_1 and the SA functions indicates that at that point the SP_1 measure probably reflects pure response bias. In terms of a response-bias interpretation of the results, the relatively high threshold of the SP_3 measure observed for high-frequency words may be taken simply as being due to the requirement of three consecutive identifications. Such a stringent criterion lowers the probability of correct chance guesses and threshold is increased. But if this were the case then the slope of the SP_3 function would have to be greater than that of either the SP_1 or the SP_3 functions.[3] Moreover, if we assume that the SP_1 threshold for the highest frequency (3.8) reflects only the probability of the corresponding responses, namely $\frac{1}{4}$, then the probability of three consecutive congruent responses would be $\frac{1}{64}$, and the expected mean SP_3 threshold would equal 64. The mean SP_3 threshold for the highest frequency, however, is only 9.1. The elevated SP_3 threshold at high frequencies must therefore be due to an increased role of the stimulus factor.

For low-frequency words the probability of a correct chance recognition is so low (1/100) that we must regard the SP_1 threshold, which at that point is significantly lower than the SA average, as also due to a stimulus effect. Since the difference in the probabilities of one correct chance identification and three consecutive correct identifications is so vast (1/100 and 1/1,00,000) and since we observe virtually no difference between the SP_1 and the SP_3 thresholds to low-frequency words, the conclusion that at low frequencies both the SP_1 and the SP_3 measures reflect the influence of perceptual processes is strengthened.

The difference between the functions obtained in the absence and in the presence of stimuli would have been considerably greater were it not necessary to restrict the series to 20 trials and thus to employ the arbitrary criterion of 21 when *S*s failed to give a congruent response before the end of the series. The assignment of arbitrary scores was necessary primarily for the SA measure. The data of the SP_1 measure include only 3% and of the SP_3 measure only 5% of arbitrary scores. However, as in the Goldiamond-Hawkins study, the proportion of arbitrary scores in the SA data was quite large. Moreover, as in their study, it decreased with increasing training frequency. Thus, for words of frequency of 1 there was 85% arbitrary scores, and for words of frequency of 25, 10%. If the recognition trials extended over several hundred trials, the SA function would show a considerably steeper slope than the one observed in the present or in the Goldiamond-Hawkins studies, while the SP_1 and the SP_3 data would remain virtually the same.

---

[3] Given $p_a > p_b$, then $p_a/p_a^3 < p_b/p_b^3$, where p_a and p_b are probability values.

Effect of motivation on response emission.—The recognition situation involves the competition of response tendencies acquired in training. According to the evidence previously cited, when responses are in competition, the probability of highly trained responses increases with increased motivation, and the probability of those less trained decreases. To the extent that response evocation in the recognition series depends not on the stimuli presented but on the strength of the habits underlying them, we should find that motivation interacts with training frequency. However, if response emission depends on perceptual processes and is independent of the strength of those habits, such an interaction should not be found. We shall first consider the effects of motivation on response emission, ignoring for the moment whether these responses represent correct recognition or pseudo recognition.

Figure 2 shows the effects of motivation on the probability of responses emitted during the recognition series. We first note for the SA condition that previous training did in fact result in a differential probability of response emission. The effect of training frequency was significant at the .001 level (see Table 1). As in the Goldiamond-Hawkins experiment, the response proportions correspond very

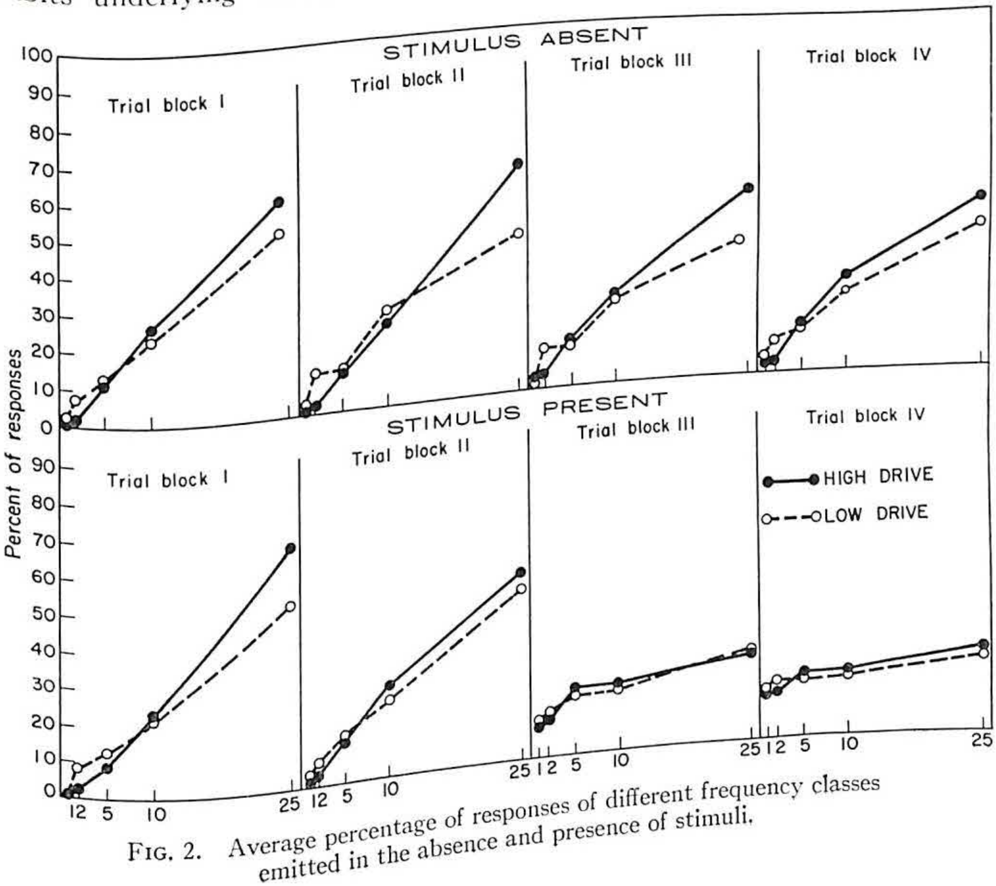


FIG. 2. Average percentage of responses of different frequency classes emitted in the absence and presence of stimuli.

TABLE 1

Analysis of Variance of the Average Percentage of Responses Emitted in the Absence and Presence of the Stimulus

Source	*df*	Stimulus Absent		Stimulus Present	
		MS	*F*	*MS*	*F*
Drive (D)	1	51.35	1.86	7.01	<1.00
Error (b)	22	27.55		28.78	
Trial blocks (T)	3	8.11	<1.00	86.38	3.18*
Frequency (F)	4	8,802.47	158.83***	3,940.85	35.02***
D × T	3	2.46	<1.00	3.38	<1.00
D × F	4	282.85	5.10***	74.44	2.64*
F × T	12	41.79	<1.00	65.58	2.41**
D × F × T	12	16.06	<1.00	530.96	19.21***
Error (w)	418	55.42		27.11	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

closely to the training frequency. Out of the 100 presentations in training two words each occurred once, twice, 5 times, 10 times, and 25 times. The proportions of these five classes of words are therefore .02, .04, .10, .20, and .50, respectively. The corresponding proportions of responses emitted by *S*s in the SA condition for both drive groups combined were .019, .051, .119, .244, and .495. The linear function fitted by the least-squares method was equal to $1.09x - .002$.

The differential probability of response emission is maintained throughout the entire recognition session, as shown by the nonsignificant F × T interaction. Moreover, we do find that there is a significant interaction between drive level and training frequency. As expected, with increased motivation highly trained responses increase in probability of emission and responses governed by weak habits (frequencies of 1 and 2) decrease. The analysis of variance (Table 1) further shows that this D × F interaction is maintained throughout the recognition series, as seen from the nonsignificant interaction between motivation (D), frequency (F), and trial blocks (T).

With the stimuli present a somewhat different picture emerges. The effect of previous training is now considerably weaker. We note that the weakening of the training effect is progressive over trial blocks. On the first trial block the curve is rather steep, but on each successive trial block the slope is reduced. This effect is demonstrated by the significant F × T interaction. Since during the first trial block the stimulus is not yet visible, the situation here is analogous to the stimulus-absent condition, thus allowing response bias to operate freely. We note that the results on the first trial block are identical to those found throughout all the trial blocks of the stimulus-absent condition. Thus, the increasing accessibility of the stimulus nullifies the effect of previous training on response emission. As stimulus information increases, responses become increasingly more dependent upon a perceptual process. If we were to continue the recognition series further, the response-emission curves would eventually show a slope of zero. The origin of these curves would be 20%, since each of the five frequency classes occurs in the recognition series equally often.

The D × F interaction shows a similar pattern. While still significant, it is considerably weaker than in the stimulus-absent condition. Also, the D × F interaction decreases over trial blocks, further showing a reduced dependence of responses on previously acquired habits. This effect is indicated by the significant D × F × T interaction.

Effect of motivation on recognition threshold.—The effects of motivation on the two recognition threshold measures and on the pseudo-recognition threshold are shown in Fig. 3. Since the effect of drive on response emission was found to decrease with increasing role of the stimulus factor, and since threshold depends on response emission, similar effects should obtain for threshold. We note from Fig. 3 that the SA functions of the high- and low-drive groups intersect, that the corresponding SP_1 functions converge, and that the two SP_3 functions are parallel. It was pointed out above that both stimulus and response factors are probably involved in the SP_1 measure. On the other hand, the *relatively high threshold for high frequencies* found for the SP_3 measure, together with its *shallow slope*, were taken to indicate that in the SP_3 data the response factor must be negligible. Since we consider the presence of drive-frequency interaction as another independent means of detecting the presence of response bias, the difference in this interaction between the SP_1 and the SP_3 measures is of particular importance. The D × F interaction for the SP_1 measure showed $F\,(4, 88) = 6.32$, significant at the .001 level, and for the SP_3 measure $F\,(4, 88) = .19$. The difference in the drive-frequency interaction between the SP_1 and SP_3 measures can also be seen from the slopes of the fitted linear functions. Since the high-drive and low-drive functions of the SP_3 measure are practically parallel,

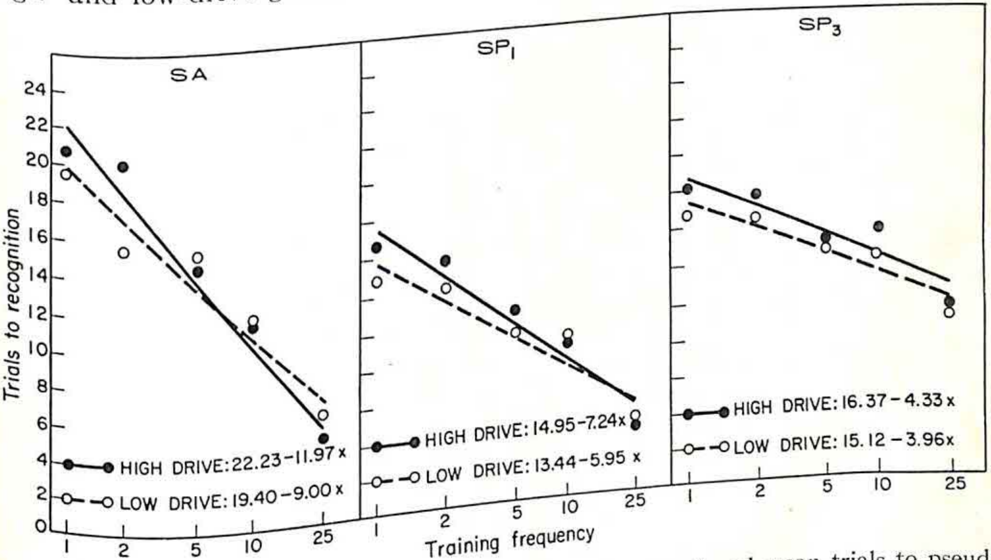


FIG. 3. Mean trials to recognition (SP_1 and SP_3 measures) and mean trials to pseudo recognition (SA measure) as a function of frequency for the high-drive and low-drive groups.

the likelihood of response bias in this measure is rather small. This finding taken together with the significant slope of the SP_3 function suggests that when threshold is assessed by the ascending method of limits and when three successive correct identifications are taken as the recognition criterion, the measure is relatively free from response bias. It is also seen from Fig. 3 that the overall threshold level for the SP_3 measure is somewhat higher under increased motivation, but this result was not significant, showing $F\ (1, 22) = 1.05$.

One apparent inconsistency of the results should be noted. It was argued that at high frequencies the SP_1 measure probably reflects pure response bias and at low frequencies a perceptual process. Thus, the effect of drive manipulation on the SP_1 measure should manifest itself primarily at high frequencies. That is, the high- and low-drive SP_1 curves should converge at low and diverge at high frequencies. The opposite, however, is the case. It must be pointed out that we are dealing here with two groups of different *S*s, and that the overall recognition performance as measured by SP_3 was consistently inferior in the high-drive group. This difference reflects in all probability a sampling difference rather than an effect of drive. If this difference between groups in the absolute SP_3 recognition level is taken into account—that is, if the overall threshold difference found between the high- and low-drive groups is either subtracted from the high-drive function of the SP_1 measure or added to the low-drive function of the SP_1 measure—the apparent inconsistency disappears.

DISCUSSION

The comparison of slopes of the three measures, the examination of response-emission data, and the effects of drive manipulation, all gave strong indications that the role of response bias in the frequency-recognition relationship has been overestimated by studies using pseudo-recognition procedures. Consider the following argument. Suppose that prior perceptual experience has no effect on threshold. Then the SP_1 measure would, like the SA measure, represent response bias alone. But since the slope of the former measure is about $\frac{2}{3}$ of the latter, we must conclude that the presence of the stimulus in the recognition series reduced the influence of response bias by about $\frac{1}{3}$.

The results of manipulating motivation showed that the SP_3 measure is very likely free of response bias, and that it probably reflects a stimulus or perceptual factor alone, while the SP_1 measure involves both factors. If this conclusion is correct, then the relative role of the response factor in the SP_1 measure may be estimated somewhat differently. We have:

Stimulus factor alone (SP_3):

$$15.78 - 4.21x.$$

Response and stimulus factors together (SP_1): $14.27 - 6.79x.$

Therefore, response factor in the SP_1 measure is equal to

$$(14.27 - 6.79x) - (15.78 - 4.21x) = -1.51 - 2.58x.$$

Comparing the estimated response factor in the SP_1 measure $(-2.58x)$ with that observed for the SA measure $(-10.49x)$, we note that the response factor loses almost $\frac{3}{4}$ of its effect when stimuli are present. Thus, the most conservative assumption leads us to conclude that the presence of the stimulus reduces the effect of response bias on the frequency-recognition slope by $\frac{1}{3}$, and the least conservative assumption that it reduces it by $\frac{3}{4}$. Necessarily, the introduction of a stringent recognition criterion reduces it further.

The failure to obtain a drive-frequency interaction for the SP_3 measure makes the presence of a response bias unlikely when a stringent recognition criterion is employed. In one previous study, however, a drive-frequency interaction was found for recognition threshold (Glucksberg, 1962). But *E*'s interest in that study was primarily in the effect of drive on response evocation, and his procedure, therefore, maximized the operation of response bias. First, the tachistoscopic presentation involved showing two words simultaneously, one of low and one of high frequency. The high-frequency word was printed in black upper-case letters, while the low-frequency word in pale red, lower-case letters. One group was instructed to identify the black and another the pale red words. Because the black upper-case words were so easy to recognize (Glucksberg, 1962, p. 38), strong response competition was created for the group instructed to report the pale red words. Secondly, and most important, the threshold criterion consisted of only one correct identification. Since a significant drive-frequency interaction was also found in the present study when this criterion was used, there is no contradiction between Glucksberg's and our findings.

While we were able to demonstrate with some certainty that the role of response bias in the frequency-recognition relationship is probably negligible when stringent recognition criteria are used, the present experiment is not capable of determining unequivocally whether the slope of the SP_3 measure is due to the prior perceptual experience with the stimuli. The problem was approached indirectly by manipulating and observing the response component of the relationship. The study of the effect of prior perceptual experience on recognition, with response bias controlled, requires either a threshold method insensitive to response effects (such as the forced-choice technique), or a design in which prior perceptual experience and prior training of response tendencies are manipulated independently of one another. Both experiments are now in progress.

REFERENCES

DUIJKER, H. C. J. *Taal en psychologisch werkelijkheid.* Amsterdam: N. V. Noord-Ned. Uitg. Mij., 1946.

FELLER, W. *An introduction to probability theory and its applications.* New York: Wiley, 1950.

GLUCKSBERG, S. The influence of strength of drive on functional fixedness and perceptual recognition. *J. exp. Psychol.*, 1962, **63**, 36–41.

GOLDIAMOND, I., & HAWKINS, W. F. Vexierversuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 1958, **56**, 457–463.

HAVENS, L. L., & FOOTE, W. E. The effect of competition on visual duration threshold and its independence of stimulus frequency. *J. exp. Psychol.*, 1963, **65**, 6–11.

SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition threshold for words. *J. exp. Psychol.*, 1952, **43**, 195–210.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

SPENCE, K. W., TAYLOR, J., & KETCHEL, R. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 306–310.

(Received May 13, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 286–295

PAIRED-ASSOCIATE LEARNING WITH MASSED AND DISTRIBUTED REPETITIONS OF ITEMS[1]

JAMES G. GREENO

Indiana University

Effects of 2 conditions of practice on individual paired-associate items were compared. Repetitions of some items were distributed in the sequence and other items were repeated on nearly successive trials. Learning was tested (a) after the massed and distributed repetitions were presented as a training sequence (n's of 24 and 36) and (b) when the massed and distributed trials were given throughout the learning of the list (n = 30). More learning occurred in the distributed than in the massed condition. The finding argues for a discrimination theory of paired-associate learning as opposed to a contiguity or a reinforcement theory, at least as regards the associative or hook-up process.

Consider two situations in which a paired-associate stimulus and its response are presented on two anticipation trials. Ordinarily, two such trials are separated by the presentation of a number of different items which *S* is learning in the same list. Now suppose that the item is presented once and then presented again immediately, or after just one other item has been presented. These two conditions can be described, respectively, as distributed and massed repetitions of an individual paired-associate item.

The question to be examined in this paper has to do with what happens on the repetitions of an item when these are massed. We know that if a stimulus-response pair is presented, and then a test trial is given immediately or shortly afterward, *S* will very likely give the correct response (Murdock, 1963; Peterson, Saltzman, Hillner, & Land, 1962; Tulving & Arbuckle, 1963). The question to be raised here is what effect massed repetitions have on *S*'s ability to give the correct response on later trials. In other words, if repetitions of an item are massed, how much learning takes place?

In the first experiment to be reported, there were three anticipation training trials for each item of a list of paired associates. For one-half of the items the three training presentations were massed; for the other items the training presentations were distributed. Then all the items were presented in a list to be learned to criterion by the standard anticipation method; these trials were treated as a test series, and *S*'s performance on the two sets of items during the test trials was compared.

In the second experiment, a list was used in which some items were presented in pairs of massed repetitions, and other items were presented in the usual distributed fashion, with these conditions applying throughout the learning of the list.

Different theoretical attitudes lead to differing expectations about the relative effectiveness of massed repetitions. In particular, from a contiguity or a reinforcement position, one would expect massed repetitions to be at least as effective as distributed

---

[1] This research was supported by Grant GB-319 from the National Science Foundation.

repetitions. However, from a discrimination or differentiation view, the distributed repetitions should be expected to have more effect.

First, suppose that contiguity between the stimulus and correct response provides the occasion for learning. The last event on every trial is the presentation of the correct response; this is true whether repetitions are massed or distributed. However, if repetitions are massed, then *S* also gives the correct response contiguously with the stimulus on a large proportion of the trials. Thus there should be at least as much learning on massed repetitions as on distributed repetitions.

Now, suppose that learning is facilitated by *E*'s giving reinforcement to *S* for making correct responses. Again, the high frequency of correct responses when tests follow earlier trials closely would be expected to give an advantage to items receiving massed repetitions so that they would be learned faster than items receiving distributed repetitions.

Quite a different expectation results if one takes the view that paired-associate learning is primarily a problem in intralist discrimination, as has been suggested by Gibson (1940), Shepard (1957, 1958), and Murdock (1958). According to this view, the problem for the learner is not so much the building up of response strength for the correct response as it is the reduction of generalization tendencies for wrong responses. When an item is presented immediately following itself, it can be assumed that generalization tendencies which would produce errors are minimal at that point. Therefore, the presentation of the item can have little or no effect in further reducing generalization. This would make the repetitions of an item relatively ineffective when the repetitions followed one another closely.[2] Putting this in less technical language, with massed repetitions presentations of an item tend to follow other presentations of the same item. With the same number of distributed presentations, presentations of an item tend to *follow presentations of different items.* The latter condition should be more effective as regards discrimination of the item from others in the list.

The task used in the present experiments was one in which no response learning or response differentiation should have been necessary. Therefore, these experiments have to do with the relative importance of contiguity or reinforcement vs. discrimination in the associative or response hook-up process in paired-associate learning.

EXPERIMENT I

Method

Subjects and design.—The *S*s were students in introductory psychology. One group of 24 *S*s received a 12-item test list containing 6 items that had received distributed repetitions and 6 items that had received massed repetitions during a training series. Then a group of 36 *S*s received a 6-item test list containing 3 items that had received training under each of the two conditions.

Materials and procedure.—The stimuli were: BATH, BREAD, COLD, DARK, EARTH, FRUIT, HARD, HOUSE, JOY, SHEEP, SLOW, STEM, KING, LONG, THIEF, and WISH. The last four were not used in the six-item condition. The stimuli were selected from those in the Minnesota norms (Russell & Jenkins, 1954). The number and frequency of associative responses which appear for more than one of the selected stimuli were kept as small as possible.

---

[2] An analogous argument about response strength is at least strained and perhaps impossible. If one says that the momentary response strength following a presentation is at maximum, then of course it cannot be increased further by an immediate repetition of the item. But then it is difficult to see why an item would not always be learned with just one presentation; or, failing this, why learning would ever occur at all.

In the 12-item condition, the responses were the digits 1, 2, 3, and 4. In the 6-item condition, the responses were the digits 1 and 2. Fewer responses than stimuli were used in order to minimize *S*'s opportunities to improve his performance on one item by eliminating responses which were paired with other items.

The stimulus words were typed on index cards and were viewed by *S* through a rectangular hole cut in a Masonite shield. The *S*s were instructed to give one of the responses on every trial. On each trial, the stimulus was in view for 3–4 sec. During this time *S* pronounced the word and gave his response, and *E* informed *S* of the correct response by saying, "—— is correct." About 2 sec. elapsed between trials, while *E* selected the next stimulus. Trials were presented without interruption throughout the session.

Arrangement of lists.—The main part of the training series consisted of a number of trials in which each of the experimental items was presented on three anticipation trials. The three repetitions of each massed item occurred on successive trials or with one presentation of another item between successive repetitions. The three repetitions of the distributed items were spaced out through the whole series of training trials. At least four different items were presented between successive repetitions of a given distributed item.[3] The average number of item presentations separating the repetitions of distributed items was 15.08 in the 12-item condition, and 8.83 in the 6-item condition. The average number of item presentations separating the repetitions of massed items was 0.25.

In addition to the experimental items which were presented in the above-described fashion, extra items were used to provide warm-up trials, to aid in satisfying the spacing requirements among the training presentations of the experimental items, and to provide a small number of "buffer" trials between the training trials and the test series. In the 12-item condition, there were 4 extra items. These were used on 12 warm-up trials at the beginning of the session, 12 trials inserted among the training presentations of the experimental items to aid in spacing, and 4 trials at the end of the training series and before the test trials. In the 6-item condition, 3 extra items were used to provide 9 warm-up trials and 9 trials inserted among the experimental training presentations. There were six item presentations between the training series and the test trials, including 3 extra items which had not been used previously.

Four different training schedules were constructed for the 12-item conditions, and six schedules were constructed for the 6-item condition. Schedules were balanced so that massed and distributed items occurred about equally often in the different serial positions of the training series. For each condition, the schedules constituted the experimental treatments in a randomized block design with six replications. A different random assignment of stimulus words as massed, distributed, and extra items was obtained for each block of *S*s. A different random assignment of responses to stimuli was obtained for each *S*, with the restriction that each response occurred at least once and not more than twice for the set of massed items and for the set of distributed items.

Following the training sequence, the experimental items were presented in randomly permuted cycles, obtained for each *S*, restricted so that repetitions of the same item were separated by at least two item presentations in the 12-item condition and by at least one item presentation in the 6-item condition. These trials were continued until *S* gave only correct responses on three successive cycles of the 12-item list or five successive cycles of the 6-item list.

Results

A summary of the data from Exp. I is given in Table 1. Note that in the first column, there were more correct responses on the second and third training presentations for the massed than for the distributed items. This simply indicates that the spacing of repetitions was successful in producing differences in the frequency of errors.

The figures in the second column are the proportions of items for which there were no errors on the test trials during which *S* was learning the list to criterion. There were more distributed items than massed items on which there was perfect performance after training. This is consistent with a conclusion that more learning took

---

[3] Due to an oversight in constructing the schedules for the 12-item condition, one of the distributed items in one of the four schedules received its second and third training presentations separated by just one item.

TABLE 1

DATA FROM EXP. I

Cond.	Proportions of Correct Responses on Training Trials 2 and 3	Proportions of Items with No Errors on Test Trials	Proportions of Items with No Errors after First Training Trial	Mean Number of Errors on Test Trials Excluding Items with Zero Errors
12 Items				
Distributed	.465	.472	.201	2.86
Massed	.924	.340	.319	2.75
90% confidence interval for difference: distributed minus massed	−.459 ± .068	.132 ± .083	−.118 ± .076	0.11 ± 0.39
6 Items				
Distributed	.704	.574	.407	2.32
Massed	.931	.454	.398	2.64
90% confidence interval for difference: distributed minus massed	−.227 ± .157	.120 ± .092	.009 ± .112	−0.32 ± 0.57

place on the distributed repetitions than on the massed repetitions.

The figures in the third column are of interest anent whether there was any learning on the second and third repetitions of the massed items. Consider what would happen if the massed repetitions had no effect. Of course, some items would be learned when they were presented for the first time, but there should be no difference between the two conditions in the effect of the first presentation. Now, if an item were not learned on the first presentation, and was given two ineffective massed repetitions, then at the beginning of the test series it would be equivalent to a distributed item immediately following its first presentation. However, on the repetitions which we are considering as having no effect, there were very few errors. Therefore, we should expect about the same number of items with no errors after the first presentation in the distributed as in the massed set, if the second and third massed repetitions had no effect. To the extent that learning occurred on the second and third massed repetitions, the high frequency of correct responses on those presentations would result in a higher proportion of massed than distributed items with perfect performance after the first presentation. The lack of a difference between the proportions in the 6-item condition, then, is consistent with a conclusion that no learning occurred on the second and third massed repetitions in the 6-item condition. In the 12-item condition, there is evidence that some learning did take place on the second and third massed repetitions; of course, this does not detract from the main finding that the distributed repetitions had more effect than the massed repetitions in both conditions.

The final result reported in Table 1 involves only those items for which there was at least one error during the test series; the result is relevant to the all-or-none hypothesis. Suppose that the items in this situation were learned in an all-or-none fashion. If this were true, then the appropriate description of the difference between massed and distributed repetitions would be that more items were learned when repetitions were distributed.

The alternative to this would be to conclude that the increments in response strength produced by the massed repetitions were smaller than when repetitions were distributed. In other words, the all-or-none hypothesis leads one to talk about the number of learned items; the incremental-strength hypothesis leads one to talk about the amount of learning on each item.

Now, according to the all-or-none hypothesis, if there is an error on any item, we know that that item has not yet been learned. It is assumed that all unlearned items are equivalent. Thus, if we calculate the mean number of errors on those items for which there were one or more errors in the test series, we should find no difference between the items which received massed and distributed repetitions in training.

However, according to the incremental-strength hypothesis, the distributed items received larger increments of strength during the training. Therefore, the distributed items which are below threshold (and hence have one or more errors) should be closer to threshold than the massed items which are below threshold. As a result, the analytic technique of removing the items with no errors should not be expected to produce equivalent sets of items. The mean number of errors in the test series should still be smaller for the distributed items than for the massed items, even when the items with no errors are left out of the calculation.

In order to avoid *S*-selection biases in calculating the means for the number of errors per imperfect item, a score was obtained for each *S*. The number of errors made on distributed items in the test series was divided by the number of distributed items with at least one error. Similarly, the number of errors made on massed items in the test series was divided by the number of massed items with at least one error. If an *S* made zero errors either on all of the distributed items or on all of the massed items, that *S* was excluded from the analysis. The means reported in the last column of Table 1 are based on 22 *S*s from the 12-item condition and 20 *S*s from the 6-item condition. The lack of substantial difference between the means is consistent with the conclusion that at the beginning of the test series, unlearned distributed items were not closer to threshold than unlearned massed items. This conclusion, in turn, is consistent with the all-or-none hypothesis.

The intervals in the third and sixth rows of Table 1 are to be interpreted as estimated bounds on the differences between the distributed and massed items. For each interval, we can be 90% confident that the true difference between the distributed and massed items is somewhere in the interval. The estimates of σ^2 used to estimate the confidence intervals in the first three columns were the residual mean-square terms of the analyses of variance for randomized blocks. The scores which were used in these analyses were the difference scores obtained for the individual *S*s. Since some of the *S*s did not contribute scores to the means in the fourth column, an analysis of variance of these scores which took schedules and blocks into account would be very cumbersome. Therefore, the estimated values of σ^2 for the confidence intervals in Column 4 were obtained by treating the difference scores for *S*s as independent measurements; that is, blocks and schedules were ignored.

The main finding of Exp. I, then, is that more items were learned when repetitions in training were distributed than when they were massed. However, part of the difference could have been due to differences involved in the order of presenting the items. First, consider the repetitions of a distributed item during the training series. One of these presentations occurred during each one-third of the training series; therefore, each of the distributed items was presented once during the last one-third of the training sequence. Now, compare this with the massed items. Each massed item received all of its training presentations within one of the one-thirds of the training sequence. Therefore, only one-third of the massed items was presented during the trials which immediately preceded the test series; all of the distributed items were presented during these trials. Could the advantage of the distributed items have been due to their greater average proximity to the test? This is possible; and a comparison among the massed items depending on their order of presentation makes it plausible. In the 12-item condition, the proportions of massed items with no errors in the test were .27, .29, and .46 for the first 2, the second 2, and the last 2 items presented during training. In the 6-item condition, the proportions of massed items with no errors in training were .39, .36, and .61 for the first, second, and third items in order of their presentation.

Experiment II

The purpose of the second experiment was to remove the possibility of a difference between distributed and massed items produced by differences in presentation orders. The massed and distributed repetitions occurred throughout the learning of the list. That is, each time a massed item was presented, it was presented a second time either as the very next item or following just one other item. The distributed items were presented on trials which were spaced out in the usual fashion. Thus, the effect of differential proximity between the training trials and the test series in Exp. I could not occur: the training and test procedures were simply combined.

Method

Subjects, materials, and procedure.—The 30 *S*s were students in introductory psychology. Three *S*s were discarded, 2 because of procedural errors by *E*, and 1 because of failure to complete the task before having to leave for a class.

The stimuli were 15 of the words listed above under Exp. I; KING was not used. The responses were the digits 1, 2, 3, 4, and 5. Instructions and method of presenting items were the same as in Exp. I.

Arrangement of lists.—Twelve of the 15 stimulus words were experimental words, and were presented in four different ways. These four conditions can be understood most easily by considering blocks consisting of 30 item presentations, and half blocks consisting of 15 item presentations. Every item appeared at least once in each full block.

First, consider two sets of items which will be called Single items. Each Distributed-Single item was presented just once in each 30-trial block. Each Massed-Single item was presented twice in each 30-trial block, with either zero or one item between the two presentations.

The other two sets of experimental items will be called Double items. These sets correspond to the above-described Single items, except that Double items were presented within each 15-trial half block. Each Distributed-Double item was presented once in each 15-trial half block; each Massed-Double item was presented twice in each 15-trial half block, with either zero or one item between the two presentations.

For items appearing near the end of a block or half block, the sequences were restricted so that at least four different items were presented before the item appeared in the next block or half block.

The three stimuli which were not in these conditions were used as extra items. They were used to provide five trials at the begin-

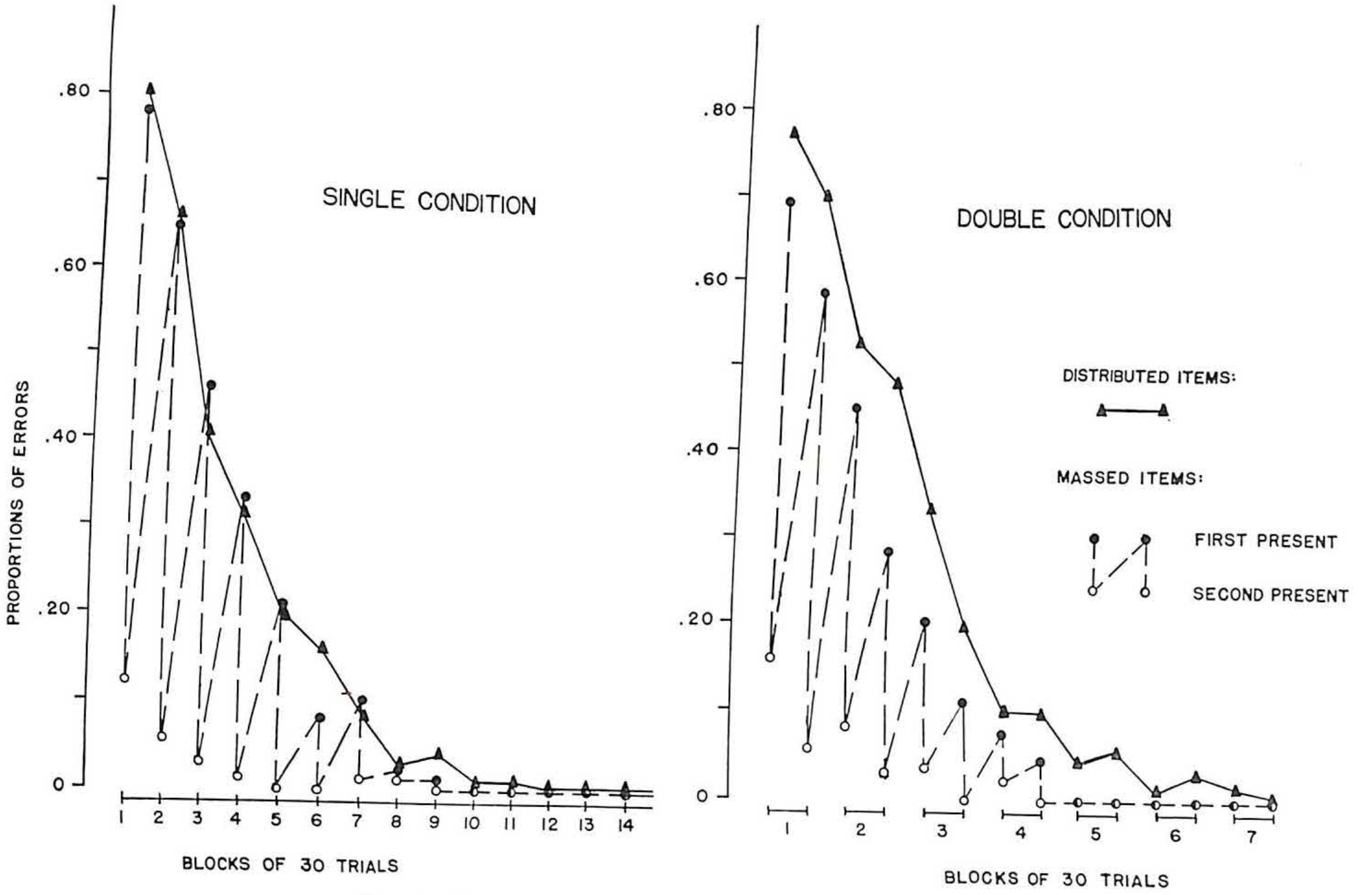


FIG. 1. Proportions of errors by 30-trial blocks in Exp. II.

ning of the experiment, in order to reduce primacy effects. Each extra item appeared once per 30-trial block during the remainder of the experiment, in positions which aided in satisfying the spacing requirements described above.

Six different schedules were prepared, and these constituted the experimental treatments of a randomized block design with five replications. A different assignment of stimuli to conditions was obtained for each block of six *S*s; there were three words in each condition, and each stimulus word was assigned to a different condition in each of the blocks. A different assignment of responses to stimuli was made for each *S*, restricted so that the three stimuli in any one condition had different responses, and so that each response was paired with at least two but at most four of the stimuli in the 15-item list. The items were presented until *S* gave only correct responses on two consecutive 30-trial blocks, excluding errors on extra items.

Results

The criterion used for stopping the experiment required different numbers of correct responses for items in the different conditions. In order to obtain fair comparisons among the sets of items, comparable criteria were applied to the items in the four experimental conditions, and trials after the criterion had been met for the set of items were excluded from the analysis. For the distributed conditions, the criterion was two consecutive presentations of all three items in which there were no errors. For the massed conditions, the criterion was two consecutive presentations of massed trial pairs on all three items on which *S* gave correct responses on the six trials that were the first members of the trial pairs.

The proportions of errors on the items in the four conditions are graphed by 30-trial blocks in Fig. 1. Each data point represents a single presentation of each item; therefore, the numbers of data points per block differ among the conditions. Perhaps the most striking aspect of the results is the lack of substantial difference between the massed and distributed items, if the second presentations of the massed items are ignored. If there were no learning at all on these second presentations, then the curve for the distributed items would be expected to fall along the upper points of the curve for the massed items. In the single conditions, this seems to have been the case, and the difference between the two curves in the double condition was not overwhelming.

Pertinent statistics are presented in Table 2. The results in the first column are relevant to the question whether massed repetitions are less

TABLE 2

DATA FROM EXP. II

Cond.	Mean Number of Presentations to Last Error before Criterion	Mean Number of Errors before Criterion
Single		
Distributed	3.38	2.74
Massed	5.50	2.88
90% confidence interval for difference: distributed minus massed	−2.12 ± 0.84	−0.13 ± 0.43
Double		
Distributed	4.38	3.40
Massed	5.41	2.84
90% confidence interval for difference: distributed minus massed	−1.03 ± 1.01	0.57 ± 0.52

effective than distributed repetitions. If each repetition of a massed item was just as effective as each repetition of a distributed item, then there would be at least as many presentations before learning for the distributed items as for the massed items. The number of presentations up to and including the last precriterion error was counted for each item; the means are presented in the first column of Table 2. The fact that a greater number of presentations was required for the massed conditions then for the distributed conditions provides further evidence that massed repetitions are less effective than distributed repetitions.

The figures in the second column relate to the question of whether the second repetitions in the massed trial pairs had any effect at all. Consider the following hypothesis: the second repetitions of massed pairs have an effect only when *S* fails to attend to the first presentation. If the hypothesis were true, then the occasions on which the second presentations could be effective would approximately correspond to the few occasions on which there was an error on the second presentation of a massed trial pair. It would be expected, then, that the mean number of errors before criterion would be the same for the distributed as for the massed items. From the second column of Table 2, it can be seen that this equality was obtained in the case of the single items, although there were more errors for the distributed items in the double condition. Thus, there is evidence that some learning occurred on the massed repetitions in the double condition, although the result from the single condition is consistent with the conclusion that no learning occurred when an item was immediately repeated.

DISCUSSION

The consistent finding of these experiments is that less learning occurs when repetitions of an item are massed than when they are distributed. In some cases there was no evidence that any learning occurred when an item was repeated very soon after a previous presentation. For reasons presented in the introduction of this article, this finding is generally consistent with a view of paired-associate learning emphasizing the importance of differentiation or discrimination among the items. The finding seems to argue against the importance of contiguity or of reinforcement in the situations studied here.

This theoretical point can be reinforced by considering the present experiments in relation to other studies in which spacing of items has been systematically varied. In two paired-associate experiments, it has been found that when similar items are grouped together in the list, learning occurs more rapidly than when similar items are spaced apart (Gagné, 1950; Rotberg & Woolman, 1963). It would be reasonable to assume that in these studies, much of *S*'s difficulty in learning the task comes from having to discriminate among the items which are similar to one another. This assumption would lead to the interpretation that when relatively indiscriminable stimuli are present in adjacent positions in the sequence of trials, then discriminations among them are more likely to occur. Then the results reported in this paper would indicate that the immediate repetition of an item seldom provides an additional opportunity for that item to be discriminated from others in the list.

REFERENCES

GAGNÉ, R. M. The effect of sequence of presentation of similar items on the learning of paired associates. *J. exp. Psychol.*, 1950, **40**, 61–73.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

MURDOCK, B. B., JR. Intralist generalization in paired-associate learning. *Psychol. Rev.*, 1958, **65**, 306–314.

MURDOCK, B. B., JR. Short-term retention of single paired associates. *J. exp. Psychol.*, 1963, **65**, 433–443.

PETERSON, L. R., SALTZMAN, D., HILLNER, K., & LAND, V. Recency and frequency in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 396–403.

ROTBERG, I. C., & WOOLMAN, M. Verbal paired-associate learning as a function of grouping similar stimuli or responses. *J. exp. Psychol.*, 1963, **65**, 47–51.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8-onr-66216, Office of Naval Research.

SHEPARD, R. N. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 1957, **22**, 325–345.

SHEPARD, R. N. Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychol. Rev.*, 1958, **65**, 242–256.

TULVING, E., & ARBUCKLE, T. Y. Sources of intratrial interference in immediate recall of paired-associates. *J. verbal Learn. verbal Behav.*, 1963, **1**, 321–334.

(Received May 17, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 296–298

SUPPLEMENTARY REPORTS

EFFECTS OF A REDUNDANT PREFIX ON IMMEDIATE RECALL

KENT M. DALLETT

University of California, Los Angeles

Ss were presented with strings of 7 or 8 digits for immediate recall. 5 conditions were compared: 7 random digits, 8 random digits, 7 digits to which S was required to add a redundant 0 prefix in recall, 8 digits including a redundant 0 prefix, and 8 digits including a redundant prefix which S was required to omit in recall. The prefix was redundant in the sense that S was advised in advance of its presence and advised that it would appear with 10 successive digit strings. Recall of 7 random digits was superior to all other conditions, which did not differ among themselves.

Conrad (1958, 1960) has shown that immediate recall of 7- and 8-digit messages is markedly reduced when *S* must prefix a zero to the string of digits in recall. This finding is somewhat surprising, since *S* always knew in advance that a zero prefix would be required, so that little or no extra learning was involved. The present experiment was undertaken in order to replicate Conrad's basic finding and to compare the effects of several zero-prefix conditions with the effect of augmenting the message by one nonredundant digit. One would expect that the addition of a zero prefix would reduce recall less than would requiring *S* to remember an extra digit. Conditions were also included in which zero prefixes were a part of the series presented to *S*.

Method.—The materials were strings of 7 and 8 digits (excluding zero) taken from a table of random permutations. These digit series were recorded on magnetic tape at a rate of approximately 2 digits per second. Three types of digit strings occurred on the tape: 7-digit series, 7-digit series prefixed by zero on the tape, and 8-digit series. Each type of digit string occurred in a block of 10 similar strings, and in the experiment, *S* was always presented with 10 strings of the same type before being switched to another experimental condition. Each string was preceded by a ready signal.

The *S* sat facing a loudspeaker, and was instructed to listen to each string as it was presented, and then, with no signal from *E*, to try to repeat the string when it was completed. The first 40 digit strings presented were alternating blocks of 10 7-digit and 10 8-digit series. Then four experimental conditions were presented, each represented by a block of 10 strings, and finally, another 40 strings made up of 7- and 8-digit series. Before each block of 10 strings, *S* was told (*a*) what type of strings were to be presented, including the nature of the repeated zero prefix if applicable, and (*b*) what *S* was to recall. The alternating blocks of 10 7-digit and 10 8-digit strings at the beginning and end of the experiment began with 7 digits for half the *S*s, and with 8 digits for the other half. The four experimental blocks were presented to each *S* in a different order, with all 24 permutations being used.

The *S*s were 24 students from the introductory psychology course at the University of California, Los Angeles. They signed up for the experiment as a part of a class requirement, and were assigned to one permutation of conditions in the order of their appearance.

The four experimental conditions could be considered as a 2 × 2 factorial design, one variable being whether or not *S* was required to prefix a 7-digit string with a zero in recall, and the other variable involving the presence or absence of a zero prefix on the tape. Thus, one of the four conditions (7:07) duplicated Conrad's main experimental condition in that no zero was presented, but *S* was required to prefix one in recall. Other conditions included a 7-digit control condition (7:7), a condition in which zero was presented on the tape and preserved in recall (07:07), and a fourth condition in which a zero prefix was presented on the tape, but *S* was told to omit it in recall (07:7).

Results and discussion.—Each digit string was scored in an all-or-none fashion: either it was completely correct or it was not. The score for each condition represented the number of correct strings in a block of 10 strings. No tendency for *S*s to improve within blocks was noted. The means and *SD*s for each condition are presented in Table 1. An analysis of variance indicated that the procedural variation of having half of the *S*s' pretests begin with 7-digit series and half with 8-digit series was not significant at the .05 level, and did not interact with any other variable. Hence, all of the data were combined, yielding 12 conditions: the 2 7-digit and 2 8-digit series of the pretest; 4 similar series from the posttest; and the 4 experimental conditions. A multiple-comparison test described by Snedecor (1957, p. 251) indicated that a difference of 1.469 was required for any 2 of the 12 conditions to differ significantly from one another at the .05 level. None of the 7-digit series differ from one another by this criterion, nor do any of the 8-digit series differ from one another. The difference between the 7- and 8-digit series is significant. All of the 3 conditions involving a zero prefix (in presentation, in recall, or in both presentation and recall), differ from the 7-digit series, but not from one another and not from the 8-digit series. Hence, the effect of the 3 zero-prefix conditions appears to be no different from the effect of adding another, nonredundant digit.

While these findings offer a striking confirmation of the zero-prefix effect, it is somewhat puzzling that the 07:7 condition did not significantly exceed the 07:07 condition in performance, since it is known that a partial report is likely to be more accurate than a full report (Anderson, 1960; Sperling, 1960). Using an individual *t* test (which capitalizes on chance), this difference would attain significance at the .05 level only if a one-tailed test were employed, $t\ (23) = 1.871$. The effect thus seems to be a small one which this experiment is not sensitive enough to detect. An attempt was made to increase the sensitivity by scoring partially correct digit strings, but there are few unambiguous ways of doing so when *S*s recall orally. The most promising approach to such scoring consisted of obtaining initial and terminal spans, i.e., the number of consecutive correct digits beginning and ending a given digit string. With series as short as these, initial and terminal spans are not independent of one another, nor are they independent of correct recalls, so it is probably not surprising that this span analysis yielded no new information.

TABLE 1

MEANS AND *SD*s OF CORRECT STRINGS PER BLOCK OF 10 STRINGS

Cond.	Number of Blocks	*M*	*SD*
7:7	5	7.27	1.95–2.97
8:8	4	4.87	2.55–2.97
7:07	1	5.08	2.83
07:7	1	5.77	2.90
07:07	1	4.69	2.70

These findings suggest that the decrement in recall resulting from the extension of a 7-digit to an 8-digit series might *not* be due to requiring *S* to remember an additional digit. However, there are several possible explanations to consider. Conrad (1960) described the addition of a zero prefix as a means of delaying recall, but also suggested that it might better be described as an interruption of recall. If delay or interruption were involved, then we might expect the 07:7 condition not to be affected by the zero prefix, which would merely serve as an extra ready signal. However, the reports of some of the *S*s suggest that it is impossible for *S* not to hear the zero as a part of the series, and that a brief interruption or delay might be involved when *S* is required to *omit* the zero in recall. A very careful latency measure might indicate that omission does delay or interrupt recall. A different possibility is that the redundant prefix adds very little to *S*'s "memory load," but acts adversely on recall by making all series more similar to one another. It would then be expected that the effect would increase gradually through a block of zero-prefix trials, which does not appear to be the case in these data. A third possibility is that the addition of a zero prefix somehow changes the manner in which *S*s record or encode the digit strings. However, such an explanation would apply only to those conditions in which a prefix was presented on the tape, and not to the condition in which *S* was required to add a prefix. This apparently minor phenomenon could be the occasion for a good deal of research.

Together with other recent studies (e.g., Peterson & Peterson, 1959), the zero-prefix effect suggests that immediate recall is extremely susceptible to interference, and that the nature of the operations capable of interfering with short-term recall is not very well understood. It would be hard, for example, to explain these findings in terms of prior or interpolated learning. More promising, perhaps, would be an attempt to determine whether there are common stimulus or response properties involved in those events which reduce immediate recall. Perhaps, for example, any stimulus capable of producing "arousal," "attention," or an "orienting response" will also interfere with recall.

REFERENCES

Anderson, N. S. Poststimulus cuing in immediate memory. *J. exp. Psychol.*, 1960, **60**, 216–221.

Conrad, R. Accuracy of recall using keyset and telephone dial, and the effect of a prefix digit. *J. appl. Psychol.*, 1958, **42**, 285–288.

Conrad, R. Very brief delay of immediate recall. *Quart. J. exp. Psychol.*, 1960, **12**, 45–47.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

Snedecor, G. W. *Statistical methods.* Ames: Iowa State Coll. Press, 1957.

Sperling, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74**(11, Whole No. 498).

(Received March 29, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 298–299

VERBAL PAIRED-ASSOCIATE LEARNING AS A FUNCTION OF GROUPING SIMILAR STIMULI OR RESPONSES[1]

IRIS C. ROTBERG

Human Resources Research Office, George Washington University

Previous experiments (Rotberg & Woolman, 1963), in which similar and dissimilar stimulus groupings were compared, indicated the superiority of similar stimulus grouping. In those experiments, the similarity categories were clearly isolated during learning. In the present experiment, procedures were employed that provided a less marked separation of the similarity categories. Although the results confirmed the findings of the previous experiments in certain respects, similar stimulus grouping was not superior to dissimilar grouping. It is hypothesized that the superiority of similar stimulus grouping depends on the functional isolation of similarity categories.

Three previous experiments (Rotberg & Woolman, 1963) indicated the superiority of a procedure that groups similar rather than dissimilar stimuli for learning. The present study compared the same variables under experimental procedures that provided a less marked separation of the similarity categories.

Method.—The experimental treatments were the same as those used in the previous study. The word list corresponded to the one used in Exp. I and II of that study. The exposure time corresponded to Exp. II and III. The method of presentation differed, as follows:

In the previous study, the following procedure was employed: (*a*) There were five learning trials, (*b*) the trials for each three-word group were completed before the next group was presented, (*c*) the test was given after all learning trials had been completed. Under this procedure, the similarity categories during learning were clearly separated.

In the present study, there were 10 trials. Each trial included the following periods: (*a*) Learning. The nine paired associates were presented. Either similar or dissimilar items were grouped. The *S*s wrote the response they considered correct after the presentation of each stimulus. They were then given the correct responses. The *S*s did not respond during the first learning trial. (*b*) Test 1. Stimuli were presented in random order. The *S*s were not informed of correct answers. (*c*) Test 2. Responses were presented randomly and *S*s were required to give the matching stimuli. The *S*s were not informed of correct answers.

Contrary to the previous study, these procedures provided a less marked separation of the similarity categories.

The *S*s were male and female high school student volunteers. There were 21 *S*s in each of the four experimental groups.

Results and discussion.—Table 1 presents the mean number of correct responses for each experimental treatment during Learning, Test 1, and Test 2. The means are given for Trial 5, Trials 1–5, and Trials 1–10. The Test means for Trial 5 by itself are directly comparable to those of the previous study, since *S*s in the latter study were tested at the completion of Trial 5. For all periods, the term stimulus group refers to stimulus grouping during Learning, while the term response group refers to response grouping during Learning.

A 2 × 2 analysis of variance for each period for each trial category indicated no significant difference between the similar and dissimilar groups. This finding was incon-

---

[1] The research reported here was conducted by the author while she was employed by the Human Resources Research Office, George Washington University, operating under contract with the Department of the Army. Opinions and conclusions are those of the writer and do not necessarily represent views of the University or the Department of the Army.

TABLE 1

Mean Number of Correct Responses for Each Experimental Treatment for Each Period

Trial Category	Period	Similar Stimulus Group	Dissimilar Stimulus Group	Similar Response Group	Dissimilar Response Group
Trial 5	Learning Test 1 Test 2	5.81 6.05 4.90	5.76 6.14 5.24	5.57 5.67 6.24	4.95 5.90 6.29
Trials 1–5	Learning Test 1 Test 2	18.62 21.62 15.10	18.29 23.33 17.38	14.71 16.71 19.71	12.29 17.19 20.48
Trials 1–10	Learning Test 1 Test 2	57.57 61.95 50.52	55.28 62.52 52.43	49.48 52.48 55.71	48.05 55.19 58.62

sistent with the results of the previous study which indicated the superiority of similar stimulus grouping.

Differences in the present experiment were found between the stimulus and response subgroups and between inside and outside substitution errors during Testing which essentially paralleled those of the previous experiments. Therefore, the only contrast in Test results was the failure to find a superiority for similar stimulus grouping in the present experiment.

It is hypothesized that no difference between similar and dissimilar grouping was found in the present experiment because the similarity categories during Learning were not clearly separated. Results demonstrating this lack of separation are provided by Table 2, which presents the mean number of inside and outside substitution errors for the similar and dissimilar stimulus groups during Learning for Trials 1–5 and Trials 1–10.

The results indicate that each experimental group made both inside and outside substitution errors during Learning in the present experiment. These results are inconsistent with the findings for Exp. II of the earlier study presented in the same table. Of the three experiments in the earlier study, Exp. II was most comparable to the present experiment in that the stimulus words and exposure times were identical. The results for Exp. II indicate that the similar stimulus group made predominantly inside errors, while the dissimilar stimulus group made predominantly outside errors. This more complete isolation of the similarity categories during Learning provided a basis for advantageous use of the similarity groups.

TABLE 2

Mean Number of Inside and Outside Substitution Errors for the Similar and Dissimilar Stimulus Groups during Learning

Exp.	Substitution Errors	Stimuli	
		Similar	Dissimilar
Present Trials 1–5	Inside Outside	5.67 8.43	6.48 8.76
Present Trials 1–10	Inside Outside	8.57 10.57	8.81 13.19
Exp. II	Inside Outside	4.93 0.07	0.00 2.64

Effective isolation of the similarity categories might occur, even when the similar groups are not clearly separated experimentally, if *Ss'* responses tend to remain within the categories. Gagné (1950) found similarity grouping and Rothkopf (1958) found dissimilarity grouping optimum. The learning categories were not presented separately in either study. However, in Gagné's study, *Ss* were given the paired stimulus and response during Learning. They were not required to respond. Any implicit responses they made were akin to responses made under a prompting procedure: They were made after the presentation of the correct response. Thus, there was little probability of extra-category responses during the learning trials. In Rothkopf's study, a confirmation procedure was employed during learning. Therefore, a substantial number of learning responses outside the similarity categories was possible. The inconsistent results of these studies can be explained by the hypothesis that isolation of the similarity categories is necessary for objective similarity grouping to be functional for *S*.

REFERENCES

GAGNÉ, R. M. The effect of sequence of presentation of similar items on the learning of paired associates. *J. exp. Psychol.*, 1950, **40**, 61–73.

ROTBERG, I. C., & WOOLMAN, M. Verbal paired-associate learning as a function of grouping similar stimuli or responses. *J. exp. Psychol.*, 1963, **65**, 47–51.

ROTHKOPF, E. Z. Stimulus similarity and sequence of stimulus presentation in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 114–122.

(Received March 25, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 300

MINIMAL PAIRED-ASSOCIATE LEARNING[1]

LLOYD R. PETERSON AND KENNETH P. HILLNER
Indiana University

The interval between 2 presentations of 2 paired associates was varied in a short-term retention study. The inferiority of the pair presented 2nd was largely eliminated in a condition in which 10 sec. elapsed between pairings. An explanation in terms of attention was suggested.

Peterson and Peterson (1962) investigated recall of two and three pairs of words after intervals up to 16 sec. filled with counting behavior. A primacy effect was found to persist through several manipulations of length of retention interval, numbers of presentations, and kind of stimulus condition at time of recall. The present experiment investigated the possibility that introduction of an interval of counting between first and second presentations might reduce the primacy effect by increasing frequency of recall for the pair originally presented second.

Method.—The technique of the previous study was adapted to the present problem. Each of 30 *S*s from introductory classes was given 96 tests following 2 practice tests. Apparatus failure resulted in discarding the data of 2 of these *S*s. Two pairs of familiar words were exposed in a memory drum at a 2-sec. rate. Both members of a pair were exposed together in presentation, while the first member was exposed alone at recall. The *S* read aloud all drum exposures. In half the tests the pairs were presented twice without interruption, while in the remaining tests an interval of 10 sec. intervened between the first and second presentation of the pair tested. This interval was filled partly by *S* counting backward from a 3-digit number exposed in the drum, and partly by exposure of the other pair. In either case the sequence of presentations was ABAB for half the tests and ABBA for the remainder. Ten seconds of counting and exposure of the other pair intervened between second presentation of a pair and its attempted recall. Three different assignments of pairs to conditions were used over the group of *S*s. All conditions were equally represented in each block of 8 tests, and a random order obtained within blocks.

Results.—Table 1 indicates that with the massed presentations a strong primacy effect was found as in the previous experiment. However, in both halves of the session the introduction of counting between presentations resulted in marked increase of recall of the second presented pair. For the second half of the session the primacy effect virtually disappeared in the spaced conditions.

An analysis of variance of number correct resulted in significance at the .01 level for Spacing, $F\ (1, 27) = 25.5$; Order, $F = 23.6$; and Half, $F = 15.3$. The Spacing × Order interaction was also significant, $F\ (108) = 10.7$.

The Spaced tests were analyzed by sequence, and it was found that the ABAB sequence, in which the B pair of the first presentation was also second in the later presentation, resulted in no poorer recall for the B pair (.60) than did the ABBA sequence (.58).

TABLE 1
PROPORTIONS RECALLED

Pair Tested	First Half		Second Half	
	Massed	Spaced	Massed	Spaced
A	.54	.57	.60	.66
B	.39	.53	.45	.65

Discussion.—The present findings can be compared with other recent studies (Peterson, Wampler, Kirkpatrick, & Saltzman, 1963) in which presentations of other pairs completely filled spacing intervals as well as retention intervals. In those studies spacing of presentations improved recall at retention intervals of 8 sec. and longer. The effect of spacing may be to maximize attention. It seems plausible that where an interval of time filled with activity occurs before repetition of a pair, attention may be greater than where such is not the case.

REFERENCES

PETERSON, L. R., & PETERSON, M. J. Minimal paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 521–527.

PETERSON, L. R., WAMPLER, R., KIRKPATRICK, M., & SALTZMAN, D. Effect of spacing presentations on retention of a paired associate over short intervals. *J. exp. Psychol.*, 1963, **66**, 206–209.

(Received April 9, 1963)

---

[1] This research was supported by Grant G 12917 from the National Science Foundation to Indiana University, a grant for which the senior author is principal investigator.

Journal of Experimental Psychology
1964, Vol. 67, No. 3, 301–302

COMPOUND-STIMULUS HYPOTHESIS IN SERIAL LEARNING

ROBERT K. YOUNG AND JAMES CLARK
University of Texas

The hypothesis that a cluster of items preceding the response is the stimulus in serial learning was tested by having 40 Ss learn 1 of 2 serial lists and then learn a "cluster" list composed of groups of 3 items taken in serial order from the previous list. The results of the learning of the "cluster" list were very similar to results of experiments testing the hypothesis that the nominal stimulus is the effective stimulus in serial learning. That is, items taken from the ends of the serial list were learned faster in the cluster list than were their controls while items taken from the middle of the list were learned no faster than their controls.

Horowitz and Izawa (1963) have recently suggested that in the learning of a serial list the effective stimulus is a cluster of items preceding the response. This hypothesis, the Compound-Stimulus Hypothesis, has previously been examined by Young (1962, Exp. II) with negative results. Horowitz and Izawa suggest that Young's procedure "failed to capture the temporal nature of the serial task" and for this reason was not a proper test of the hypothesis. They further suggest that the "elements of each cluster should appear singly as they do in the serial task." The present experiment tests the Compound-Stimulus Hypothesis in the manner suggested by Horowitz and Izawa.

Method.—Each of 40 *S*s learned one of two 15-item serial lists in which the first item served as the anticipation symbol. Each *S* then learned a "cluster" list composed of 10 groups of 3 items each, taken from the 30 items of the two serial lists. Of the 10 groups of items, sublists of 5 groups were taken from each of the serial lists with the result that each *S* learned 5 clusters from the previously learned serial list and 5 new clusters. The experimental design is a 2 × 2 mixed factorial in which the columns represent the two sublists of items in the cluster list and the rows represent the two serial lists. Each of the three items in a cluster, $n-2$, $n-1$, and n, was presented alone, and the items appeared in the same order as in the serial list. Each of the $n-2$ items served as the nominal stimulus for the subsequent $n-1$ item which in turn served as the nominal stimulus for the subsequent n item.

The serial lists were presented at a 2-sec. rate with a 6-sec. intertrial interval. The cluster list was also presented at a 2-sec. rate with a 2-sec. interval between clusters and a 6-sec. intertrial interval. The items were two-syllable adjectives taken from Melton and Safier (Hilgard, 1951) and were as unrelated as careful inspection would permit. The learning criterion for each list was one perfect recitation. Four orders of cluster presentation were used to avoid serial learning. Both lists were presented by a filmstrip projector.

Results and discussion.—The first analysis is in terms of the number of trials to learn the two 15-item sublists of the cluster list. For the experimental sublist (i.e., those items previously learned in the serial list) the mean was 12.63 trials while for the control sublist it was 14.14 trials. This difference is not significant, $F(1, 38) = 2.85$. The second analysis is in terms of the number of trials to learn the experimental and control sublists of the five n items. Here the experimental and control means were 6.50 and 9.80, respectively, and the difference is highly significant, $F(1, 38) = 14.50$, $p < .001$. The third analysis is in terms of the number of trials to learn the two $n-1$ sublists. Here the experimental and control means were 11.50 and 13.83, respectively, and again the difference is significant, $F(1, 38) = 7.39$, $p < .01$.

Finally, the learning of individual items was analyzed as a function of the position they held in the serial list. The mean correct responses per trial for each n and each $n-1$ item were compared to the mean correct responses for the control item which was taken from the same serial position of the unlearned serial list. Of the 10 n and $n-1$ items, 3 were learned significantly faster than their respective control items; those which had held Positions 2, 3, and 15 ($F = 36.04$, 31.12, and 24.04) in the serial list.

The data appear to support a hypothesis which suggests that the effective stimulus in serial learning is the preceding items in the list. Such an interpretation is, however, not without difficulties. The $n - 1$ items, preceded only by a single item ($n - 2$), yielded positive transfer, suggesting that the effective stimulus is the *single* preceding item rather than a cluster. It should also be noted that the learning of items as a function of previous serial position conforms almost exactly to comparable results in previous experiments designed to test the hypothesis that associations are formed between adjacent items in a serial list (Young, 1961, 1962). Since a test of the Compound-Stimulus Hypothesis does not preclude transfer resulting from associations formed between adjacent items during serial learning, it is not surprising that differences found in previous experiments are also found in the present data. Thus the present data cannot be unambiguously interpreted as supporting the cluster hypothesis.

In addition, the analysis of the learning of the n sublists lends relatively little support to the cluster hypothesis. Two of the five items in this sublist were learned rapidly. These items (3 and 15) were taken from the ends of the serial list, and previous research has indicated that such positive transfer is found only for the ends of the list. The most parsimonious interpretation of these findings is that at the extremes of the list associations are formed between adjacent items. The lack of positive transfer for items from the middle of the serial list in this and previous experiments indicates, however, that such interitem associations are formed *only* at the extremes of the serial list.

REFERENCES

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517–567.

HOROWITZ, L. W., & IZAWA, C. Comparison of serial and paired-associate learning. *J. exp. Psychol.*, 1963, **65**, 352–361.

YOUNG, R. K. The stimulus in serial verbal learning. *Amer. J. Psychol.*, 1961, **74**, 517–528.

YOUNG, R. K. Tests of three hypotheses about the effective stimulus in serial learning. *J. exp. Psychol.*, 1962, **63**, 307–313.

(Received May 6, 1963)

Journal of
Experimental Psychology

Vol. 67, No. 4 April 1964

LONG-TERM MEMORY AS A FUNCTION OF RETENTION TIME AND REPEATED RECALLING[1]

EDWARD A. BILODEAU
Tulane University

MARSHALL B. JONES
University of Florida

AND C. MICHAEL LEVY
University of Wisconsin

This is the 4th in a series on the long-term retention of single, discrete responses. In this one, Ss were required to recall 3 initial events on 4 subsequent occasions. These events were the original response of Training (R_0), the alternative response to R_0 (R^a), and the informative feedback pertaining to the adequacy of R_0 (KR). In terms of means and variances, recall was no better or worse for the Ss repeating responses than for comparison Ss who recalled but once. 2 findings grew out of an analysis of the correlation matrix of intermemory trials: (a) the later memory trials were more related to the 1st memory trial than to the training trial and (b) a single factor was sufficient to explain the pattern of correlations for attempted repetitions of R_0 and R^a. The 2nd finding confirmed the hypothesis of Bilodeau and Levy that forgetting is a process of interference which accumulated at random.

A procedure often used in the study of memory has been patterned after one employed for the study of learning. During Recall, the procedure is characterized by *E* saying, "do your best on the next trial" and seldom is *S* asked to "repeat your first response."

Three studies of memory have been completed where the latter instruction to *S* forms the dominant theme (Bilodeau & Levy, 1964; Bilodeau, Levy, & Sulzer, 1963; Bilodeau, Sulzer, & Levy, 1962).[2] The present study differs from the others in that *S* was asked to repeat his first response not

---

[1] This work was partly financed by Grant G-13294 of the National Science Foundation and partly by the Graduate Council on Research of Tulane University. The writers are much indebted to J. L. Sulzer and to W. H. Saufley, Jr. for their assistance in the collection of data.

[2] This is the fourth of four comprehensive studies of levering in a project emphasizing those events within trials that might be subject to forgetting. Taken as a whole, the four experiments elaborate new treatments of the independent and dependent variables of memory. For a wider perspective on these four studies and their relations to other contemporary research see Bilodeau and Levy (1964).

once but four times. The repetitions will be designated 1, 2, 3, and 4; the training or original response will be designated 0. The general question can now be properly phrased: What is the effect on future recalling of earlier recalling? Recall trials are not important if R_0 is the standard against which the *n*th response is produced; they are important if recalled responses serve as the standard. Memory trials are also important if they serve as interpolated responses interfering with the trace of R_0.

Data collected by Bilodeau and Levy (1964) will be used as the control comparison for the experiment to be reported. They used different groups of *S*s for each retention interval. That is, a given *S* recollected but once; either 3 min., 20 min., 2 days, or 6 wk. were placed between R_0 and R_1. In the present study, the same intervals also separate R_1, R_2, R_3, and R_4 from R_0, respectively.

A trial of recall may serve the function of a trial of training. It is possible that in some instances of recalling, *S* recalls not so much the original event as some more recent memory of that event. That is to say, part of the variance in recalled responses is attributable to the original event (R_0) and part is attributable to later attempts at repeating the original ($\sigma_n^2 = \sigma_{n-1}^2 + \sigma_0^2$). This sort of analysis clearly provides for the wandering of memory or the growing variable error in a series of recollections of an original event.

Methodologically, it is important to distinguish between the procedure of repeated recalling (retrieving or retention testing) and of repeated practice (storing or acquisition training), the former referring to repetitions with no improvements required and the latter to a series of improvements with no repetitions required. Bilodeau and Bilodeau (1961) have reviewed a long-neglected literature on the analysis of intertrial correlations bearing on the learning process, and were unable to find a single analysis for a retrieving process. Thus it is not known whether the correlation patterns of recalling resemble those of learning. The present study can do no more than to open up what appears to be a substantial area.

In learning, the matrix of intertrial correlations always has the "superdiagonal form." A trial correlates strongest with its neighbors and weakest with the first and last trials of the series, producing a pattern with correlations highest in the superdiagonal (just above the main diagonal) and which decreases monotonically when moving either upwards or to the right. Jones (1962) has interpreted these findings as meaning that all known examples of learning represent a process of simplification. The present experiment will help decide if a similar pattern and interpretation can be made of the intercorrelations of memory trials.

METHOD

Subjects and apparatus.—The *S*s were 132 basic airmen at Keesler Air Force Base.[3] Of these, 23 were eliminated because they were absent for one or more sessions of Recall. Four *S*s did not recall knowledge of results at all and one *S* gave such a large one that he too was discarded.

The apparatus, shown in Fig. 1, is known as Manual Lever D and was described in detail by Sulzer (1963). The *S*'s task is to stand in front of the box which houses the lever, grasp the unseen lever, and move it from left to right on the command of "move."

---

[3] The *S*s were provided through the cooperation of the Commander and of Major Marvin Karp, Director of Military Training. We are especially grateful because the repeated testing of 132 *S*s over a 6-wk. period is not possible at most civilian institutions.

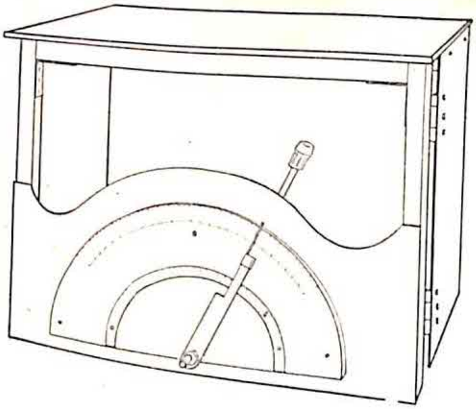


FIG. 1. Schematic view of Manual Lever D showing the lever positioned at Target 100.

The resting position of the lever was 44° above the horizontal, Target 100 was 20° beyond the starting position, and Target 300 was located 60° beyond. Responses and Targets may be converted to degrees of arc by dividing by 5.

Procedure.—The details of Pretraining and Training were described by Bilodeau, Sulzer, and Levy (1962) and so there is no need to repeat them all here. Recapitulating very briefly: during Pretraining the lever was made to stop automatically at one, then at the other, of two targets. In other words, S's hand was guided to two places. Thereafter, there was no automatic stop and S was required to stop the lever without mechanical assistance. During Training, S guessed which of the two targets was "Correct" and indicated his selection by moving the lever (R_0) in an attempt to hit the target. The KR was administered immediately afterward and E recorded the amplitude of the response.

Unlike the preceding studies where S retrieved the events but once, S was used in all four retention intervals: 3 min., 20 min., 2 days, and 6 wk. The S's major activity during the 3-min. interval related to walking into another room where Recall was administered. There he was asked for R_1, R^a_1, and KR_1. Afterwards, S was directed to a room where he waited silently with other Ss for "a similar experiment." Shortly before the 20-min. session, S was ushered back to the Training room. The three to-be-recalled events were scheduled in the same order as in the first Recall session: R_2, R^a_2, KR_2. Two days later S was again tested and R_3, R^a_3, and KR_3 were obtained; after 6 wk., R_4, R^a_4, and KR_4 were called for.

Those Ss whose R_0 was less than 200 units of amplitude were told, "That was the wrong target. You did not move far enough by ____ units." These Ss are named *Low* or *100 Shooters* and their R^as should be in the vicinity of 300. Those Ss making R_0 greater than 200 were told, "That was the wrong target. You moved too far by ____ units." They are the *High* or *300 Shooters* and their R^as should be in the vicinity of 100. No S was told that his R_0 was correct because there was a shortage of Ss for such a condition.

The reasons for administering KR as described above were elaborated earlier in the series. One reminder may be helpful. Since an aim of the inquiry was to investigate the cue properties of KR, it was necessary to avoid confounding R_0 and KR. The E achieved this by sampling rectilinearly from the series 27, 54, 81 . . . 324 in his programing of KR and assigned one value unsystematically to S irrespective of the magnitude of R_0. The KR pertains to the distance between R_0 and the alternative target though the numbers are arbitrary.

Results and Discussion

The means are not reported here because they are similar to results fully described before by Bilodeau et al. (1962). In effect, the means remain stable over 2 days of retention and then change. After 2 days, the R_3's of the 100 Shooters rise toward 300 and the R_3's of the 300 Shooters fall toward 100.

Variances.—The variances are shown in Table 1; $n = 47$ and 57 for the 100 and 300 Shooters, respectively. The 300 Shooters will not be described, their data are presented as supporting evidence.

TABLE 1

Variances for 100 and 300 Shooters

	Response				
	0	1	2	3	4
100 Shooters					
R	943	837	1136	2792	10514
R^a		1476	1357	1367	8717
KR	8462	25272	29171	31093	24062
300 Shooters					
R	1616	1738	2193	2776	6449
R^a		2277	2700	2899	9097
KR	8547	15527	14804	17049	12575

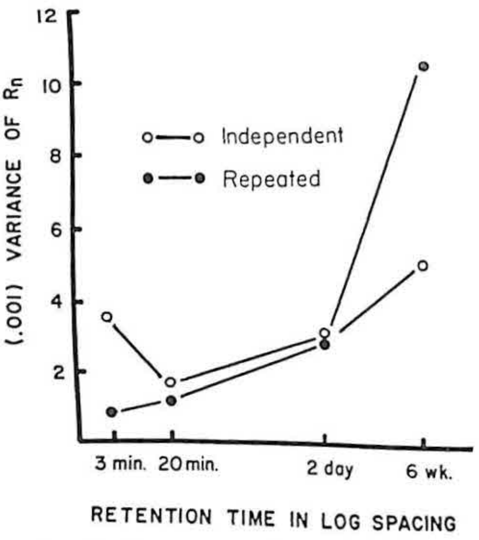


FIG. 2. Variance as a function of retention time. (The curves compare the condition of repeated recalling—from the present study—with the condition of recalling but once—from Bilodeau & Levy, 1964.)

The R_1 of 837 is a good match for the R_0 of 943, KR_1 exceeds KR_0 by a factor of 3, and the values of R^a_1, R^a_2, R^a_3 remain about the same for 2 days. After 6 wk. have elapsed, R_4 is many times greater than R_0. There are many indications of forgetting during the 6 wk. These results, however, cannot be attributed to recalling qua recalling for, as we shall see, the values of Table 1 are similar to those obtained by Bilodeau and Levy when unfilled intervals were used.

The variances for R_n have been plotted in Fig. 2 along with the comparison values of the earlier study. The present variances resemble the earlier ones in magnitude and in a strong trend of positive acceleration—although the trend of the repeated responses is more regular. As a matter of fact, the new variances look like the *idealized* function previously proposed by Bilodeau and Levy to describe the results for independent groups.

The R^as and the recalled KRs were also compared across studies and they too showed a high degree of similarity.

Correlations.—The correlations of R_0 with each of the four recall responses are about the same as reported for the independent groups of Bilodeau and Levy, though $R_0 \cdot R_2$ is low for 100 Shooters. In this study we also have, in addition to the correlations with R_0, the correlations among the four responses of Recall. The intercorrelations for the levering responses are presented in Tables 2 and 3 where they are arranged in four matrices.

The pattern in all four matrices is consistent. Excepting the correlations for R_0, *which we ignore for the time being*, each matrix is at least roughly a monotonic hierarchy; i.e., the largest correlation is between R_1 and R_2, the smallest tends to be found between R_3 and R_4, and in between, the correlations decrease as one moves across the rows or down the columns (Jones, 1960).

TABLE 2

INTERRESPONSE CORRELATIONS FOR 100 AND 300 SHOOTERS DURING REPEATED RECALL OF R_0

Measure	Responses				Loading
	R_1	R_2	R_3	R_4	
100 Shooters					
R_0	75	44	56	17	
R_1		73(72)	71(72)	22(17)	85
R_2			69(72)	08(17)	85
R_3				32(17)	85
R_4					20
300 Shooters					
R_0	61	59	41	54	
R_1		72(72)	61(60)	44(47)	85
R_2			58(60)	48(47)	85
R_3				38(39)	70
R_4					55

Note.—Theoretical values in parentheses.

TABLE 3

Interresponse Correlations for 100 and 300 Shooters during Repeated Recall of R^a

Measure	Responses				Loading
	R^a_1	R^a_2	R^a_3	R^a_4	
100 Shooters					
R^a_1		74(81)	58(63)	−37(−32)	90
R^a_2			65(63)	−29(−32)	90
R^a_3				−14(−25)	70
R^a_4					−35
300 Shooters					
R^a_1		68(72)	57(60)	46(47)	85
R^a_2			64(60)	53(47)	85
R^a_3				32(39)	70
R^a_4					55

Note.—Theoretical values in parentheses.

This pattern stands in striking contrast to the pattern universally observed for acquisition trials. When trials of practice are correlated, the correlations have always fallen into the superdiagonal form (Jones, 1962). The largest correlations are between neighboring trials; the smallest correlation is in the upper right-hand corner; and in between the correlations decrease as one moves across the rows or *up* the columns. The difference between the two patterns is most obvious in the lower right-hand corner of the matrix. In the superdiagonal forms of practice the correlation between the last two trials is almost always the largest correlation in the matrix; whereas in our results the correlation between the last two recall responses is, in absolute magnitude, the smallest in three out of four matrices. The probability of this result alone, quite apart from the rest of the pattern, is less than 1 in 200.

This difference in pattern is not superficial. Superdiagonal forms and monotonic hierarchies do not submit to the same or even similar interpretations. In the present study every one of the (four) patterns can be explained as arising out of a single common factor, and without much error, particularly in the case of repeating R_0 responses among 300 Shooters. The figures in parentheses in Table 2 are the theoretical correlations under a unifactorial interpretation. The loadings on the one common factor appear at the right-hand margin and were obtained by the procedures recommended by Jones (1962, p. 285).

The superdiagonal form cannot be explained by a single common factor. Forgetting, therefore, at least as it appears in this study, is neither the same nor the inverse of acquisition. Something quite different must be happening.

Previous evidence suggests that forgetting is brought about by interference from experiences either preceding the remembered event (proaction) or following it (retroaction). In itself this notion does not require either the monotonic or the superdiagonal form of intercorrelation. But the manner in which interference acts has definite consequences for the two patterns. If interference acts in a directed or systematic way, then the correlation pattern will be superdiagonal. If it acts in a nondirected or random way, it will be monotonic.

Suppose that at some time between R_2 and R_3 something occurs which causes *S* to undershoot. It follows that *S* will undershoot on both R_3 and R_4, provided only that the process does not dissipate entirely in the meantime. Another individual, in the same period, may be led to overshoot on R_3 and R_4. The result is an element of covariance between R_3 and R_4. If the increasing variance of the recall responses has its origins in effects of this sort, the superdiagonal form follows necessarily.[4] In other words, if interference is systematic, the closer together two responses are in time the more likely it is that undershooting or overshooting is determined by the *same* interference processes. Hence, the correlation between them will be greater. Responses which span greater intervals will more frequently be affected by different processes, some of which affect the first response but die out before the second is made, and others which do not begin until after the first response has been performed. If these interfering processes (*a*) cause *S* either to overshoot or undershoot and (*b*) act over a solid block of time, then neighboring responses will correlate more with each other than widely separated ones. And the correlations will have the superdiagonal form.

---

[4] Studies of the acquisition process generally maintain a constant intertrial interval and produce a progressive diminution of trial variance.

Since our correlations do not have the superdiagonal form, systematic interference is not responsible for the increasing variance in recall responses. Either the processes at issue come and go, i.e., are active for a while, stop, and then become active again, or they do not cause a person to overshoot or undershoot. The latter, most probably, is the case. Interference in this study does not seem to be directed. It does not cause *S* to shoot one way or the other but merely increases the variability with which he shoots. At each trial *S* tries to reproduce R$_1$, but with the accumulation of time (and interference) his efforts are attended by larger and larger amounts of random variance. The "one common factor" is his memory of R$_1$. This memory, however, is attenuated with time by increasing amounts of error. With respect to R$_1$, successive responses become more and more variable, less and less reliable. From this interpretation the monotonic hierarchy follows directly.

It is worth emphasizing that this conclusion does not depend upon our choice of retention intervals. Under the hypothesis of a single common factor the spacing of the recall responses affects the factor loadings but not the pattern of correlations they generate. In this study the factor loadings fall off slowly at first, sometimes not at all, and then drop sharply between the third and fourth recall attempts. This result is altogether as it should be because the lapse of time between the last two attempts is much the longest in the series. There is, therefore, more time for random variance to accumulate. If the responses were spaced more evenly, the fall-off in factor loadings ought also to be more regular. But the monotonic pattern would remain, and it would still submit to interpretation in terms of a single common factor.

It is also worth noting that the absence of the superdiagonal form confirms the conclusion already reached that the recall responses do not interact. If at each successive response, *S*s tried to reproduce their *last* response, rather than R$_1$, the result again would be the superdiagonal form. The correlations would be strongest between neighboring trials and weakest between remote ones.

In the matrices for KR in Table 4 something of this sort seems to have happened. These matrices, unlike those for the motor responses, tend toward the superdiagonal form. Certainly, they are not monotonic. It is not possible, therefore, to regard the correlations among the several attempts to recall KR as dominated by accumulating random variance. In Table 4 there is a definite tendency for

TABLE 4

INTERRESPONSE CORRELATIONS FOR 100 AND 300 SHOOTERS DURING REPEATED RECALL OF KR$_0$

Measure	Responses			
	KR$_1$	KR$_2$	KR$_3$	KR$_4$
100 Shooters				
KR$_0$	32	26	23	−04
KR$_1$		89	82	47
KR$_2$			92	53
KR$_3$				47
300 Shooters				
KR$_0$	52	52	50	47
KR$_1$		87	92	39
KR$_2$			94	39
KR$_3$				49

neighboring attempts at recall to correlate more strongly with each other than more separated attempts. The source of this difference may be that the KR responses are verbal. This is, however, only postponing a good answer and we shall have to wait until the difference in pattern itself becomes the object of inquiry. For the present there is no satisfactory explanation and we must be content with the observation that previous recalls of KR seem to obtrude more frequently and more importantly than they do with levering responses.

It remains to explain why R_0 does not participate in the monotonic process. In general, the correlations for R_0, although they have roughly the correct form, i.e., decreasing with time, are too low. The correlations, for example, between R_1 and R_2 are fully as large as those between R_1 and R_0, if not larger. R_2 correlates much more strongly with R_1 than it does with R_0. In fact, R_0 adds nothing to the variance of R_2 which is not already accounted for by R_1. If R_0 were the leading response in the monotonic process, none of these relationships should obtain and the responses of Recall should correlate more with R_0 than with any other response.

Bilodeau and Levy, after examining many retention curves for lever positioning data, proposed a two-stage function for the retention process, the first being very short. If this proposal is correct, the correlations for R_0 may be too low because R_0 precedes the onset of the second stage. It might also be that *S*s, for some reason, choose to model their later responses of Recall not on the training trial as demanded by *E* but on the first response of Recall. This is a particularly appealing hypothesis in view of the fact that but one trial of Training was administered and *S*s were not told of a forthcoming test of retention.

Conclusions.—The data confirm those obtained earlier in the series in the view that more forgetting takes place later than earlier. The pattern of *r*'s in the intertrial matrix suggested that, unlike the trials of training reviewed before by Jones, the intercorrelations of trials of memory for levering can be explained by a single common factor. Thus, forgetting of R_0 or R^a was neither the same nor the inverse of acquisition. Of the many reasons for the difference, a prominent one is that *S* was asked to repeat his response and was not asked to improve upon it. A good deal of work on patterns needs to be done before we shall know the circumstances producing one or some other.

The forgetting of R_0 and R^a was attributed to interference, and supported the hypothesis of Bilodeau and Levy that interference accumulated during the retention intervals in a manner similar to the accrual of the occurrence of random events over time. The forgetting of KR, on the other hand, was more systematic and was not dominated by accumulating random variance.

REFERENCES

Bilodeau, E. A., & Bilodeau, I. McD. Motor-skills learning. *Annu. Rev. Psychol.*, 1961, **12**, 243–280.

Bilodeau, E. A., & Levy, C. M. Long-term memory as a function of retention time and other conditions of training and recall. *Psychol. Rev.*, 1964, **71**, 27–41.

Bilodeau, E. A., Levy, C. M., & Sulzer, J. L. Long-term retention under conditions of artificially induced recall of related events. *Percept. mot. Skills*, 1963, **16**, 895–910. (Monogr. Suppl. No. 8-V16)

Bilodeau, E. A., Sulzer, J. L., & Levy, C. M. Theory and data on the interrelationships of three factors of memory. *Psychol. Monogr.*, 1962, **76**(20, Whole No. 539).

Jones, M. B. Molar correlational analysis. *USN Sch. Aviat. Med. Monogr.*, 1960, No. 4.

Jones, M. B. Practice as a process of simplification. *Psychol. Rev.*, 1962, **69**, 274–294.

Sulzer, J. L. Manual Lever D: A basic psychomotor apparatus for the study of feedback. *Percept. mot. Skills*, 1963, **16**, 859–862.

(Received June 17, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 310–315

BACKWARD LEARNING AND THE STIMULUS-FAMILIARIZATION INHIBITORY EFFECT[1]

SEYMORE SIMON AND GORDON WOOD[2]

Northern Illinois University

This study tested the hypothesis that backward learning during the stimulus-familiarization phase of a PA (paired-associate) list may either facilitate or hinder the learning of a criterion PA list. PA lists were constructed in such a way that if backward learning occurred during familiarization training, then the criterion PA learning of 1 group would be facilitated while the criterion learning of a 2nd group would be inhibited. Ss were 76 male college students. Results indicated that the facilitation group made fewer errors ($p < .01$) and more correct responses ($p < .001$) than did the inhibitory group on the PA test list; 2 control groups made significantly fewer correct responses than did the facilitation group ($p < .05$) and significantly more correct responses than did the inhibitory group ($p < .05$). The present data support the conclusion that backward learning during PA familiarization training may either inhibit or facilitate subsequent PA learning.

A fundamental difficulty in comparing the results of studies concerned with investigating the effects of prior stimulus familiarization on criterion paired-associate (PA) learning lies in the multiplicity of stimulus-familiarization procedures employed by various *E*s. For example, during the familiarization phase of the Cieutat (1960) study, paralogs were presented to *S*s on white cards and *S*s were simply required to look at each paralog for a 60-sec. period. Gannon and Noble (1961) asked *S*s to pronounce a list of dissyllables as each was presented on a screen during familiarization training. Underwood and Schulz (1960) asked *S*s either to spell low-association-value nonsense syllables (Exp. 1) or to learn syllables as PA responses (Exp. 4) during familiarization training. As Gannon and Noble (1961) have pointed out, until considerably more empirical information about each of these familiarization procedures, among other methodological differences, is obtained, any comparative evaluation of the results of these studies remains impossible.

The present study was designed to provide additional information about the effects of PA stimulus familiarization on criterion PA learning. In addition, one hypothesis advanced by Underwood and Schulz (1960, Exp. 4) to account for the decremental effect of prior PA stimulus familiarization on subsequent PA learning was examined. The authors hypothesized that "inhibition" occurred during the test phase because the stimulus syllables in the situation tended to elicit their backward associates from the familiarization phase which interfered with the learning of the PA test list. Although other *E*s have demonstrated that backward learning may produce either positive or negative transfer effects (McCormack, 1961; Murdock, 1958; Storms, 1958; Umemoto & Hilgard, 1961), the lack of appropriate control groups precludes generalizing the results of these latter studies to a design concerned with studying the effects of prior PA stimulus familiarization on subsequent PA learning.

---

[1] Portions of this paper were presented at the Midwestern Psychological Association Meetings in Chicago, May 1963.

[2] Now at Northwestern University.

The present study employs an experimental design which permits an adequate test of the Underwood and Schulz hypothesis. Moreover, by incorporating some of the empirical findings of mediation theory into the present design, an attempt was made to relate the PA familiarization procedure utilized by Underwood and Schulz (1960, Exp. 4) to the more "traditional" transfer-of-training paradigms and their mediational aspects. Specifically, it was hypothesized that if the response to be learned in testing was a common word associate of the stimulus used in familiarization training, then test-list learning would be facilitated. For example, if backward learning occurs during the initial learning of the syllable FOJ in the PA TABLE-FOJ, then facilitation should occur in the test-list learning of the PA FOJ-CHAIR as a result of TABLE acting as a mediator between FOJ and CHAIR. On the other hand, inhibition would be predicted when the backward association learned in familiarization training did not act as a mediator in test-list learning. Thus, if backward learning takes place during the familiarization learning of the PA TABLE-FOJ, then the dominant response elicited as an implicit association to TABLE, viz. CHAIR, would hinder the test-list learning of the PA FOJ-MUSIC.

METHOD

Design and stimulus materials.—The experimental design called for the use of six groups of *S*s, two experimental groups and four control groups. The two experimental groups, the Relevant-Mediated (RM) group and the Relevant-Nonmediated (RNM) group, and Control Groups 1 and 2, the Irrelevant-Mediated (IRM) group and the Irrelevant-Nonmediated (IRNM) group, were required to learn two lists of eight PAs. For these four groups, high-commonality words having dominant associates ranging from 675–840 (Russell & Jenkins, 1954) were used as stimuli during first-list (familiarization) learning. The response terms employed in the familiarization list were syllables selected from Noble's (1961) *m'* norms. Groups RM and RNM learned a *different* set of syllables as responses to the above stimulus words than did Groups IRM and IRNM, necessitating the construction of two different lists of eight syllables each. *M'* values ranged from .73 to 2.16 with both lists of syllables having mean *m'* values of 1.40. The first-list PAs for Groups RM and RNM were: TABLE-FOJ, HIGH-CEH, DARK-BIW, LONG-GOX, MAN-RUY, BOY-MEQ, BLACK-YIL, KING-VAK; the corresponding syllables for Groups IRM and IRNM were: QUG, XON, ZAB, LIH, SIW, PIB, JEC, KIZ. For these PA lists, as well as for the test-list PAs, all of the words had Thorndike-Lorge frequency counts of at least 100/million. In all instances, an effort was made to keep intralist and interlist similarity at a minimum.

In test-list learning the syllables learned as responses during familiarization training by Groups RM and RNM were now presented as PA stimuli for *all* groups. In this test phase the response terms for Groups RM and IRM (mediated groups) consisted of the high-commonality word associates of the stimuli used in the familiarization phase. The response terms for Groups RNM and IRNM (nonmediated groups) consisted of words not associated with the stimuli used in the familiarization phase (Russell & Jenkins, 1954). Corresponding to each of the familiarization list stimulus words reported above, the test-list response words for Groups RM and IRM were: CHAIR, LOW, LIGHT, SHORT, WOMAN, GIRL, WHITE, QUEEN; the corresponding test-list response words for Groups RNM and IRNM were: MUSIC, RIVER, SWEET, FRUIT, HEAD, JOY, GREEN, HEAVY. Control Groups 3 and 4 received no familiarization training prior to learning the two PA test lists, respectively, i.e., Group C-3 learned the same test list as did Groups RM and IRM, while Group C-4 learned the same test list as did Groups RNM and IRNM. The purpose of Groups C-3 and C-4 was twofold: to determine whether there were differences in the difficulty of the two criterion lists and to determine the performance of groups receiving no prior training on a PA familiarization list.

Subjects.—Eighty-four male, Northern Illinois University undergraduates enrolled in introductory psychology courses were randomly assigned to six groups with the only stipulation that there be 17 *S*s in each of the first four groups and 8 *S*s each in Groups 5 and 6.

Procedure.—Following standard PA learning instructions each *S* learned a practice list of three PAs to a criterion of two errorless trials. The nouns HORSE, ROSE, and IRON, and the syllables PIC, LUR, and TEJ constituted, respectively, the stimulus and response terms for the practice trials. Immediately following the practice list the familiarization list was learned to a criterion of three consecutive errorless trials. The *E* engaged *S* in casual conversation for approximately 2 min. between the familiarization and test lists. Criterion for the test list was 3 consecutive errorless trials or a maximum of 30 trials. The *S*s who did not learn the familiarization list in 35 trials or less were dropped from the experiment. A total of 8 *S*s was discarded for failing to reach criterion on the familiarization list with 2 *S*s being lost from each of the groups. There remained 15 *S*s in each of the four familiarization groups, with a total experimental population of 76.

All lists were presented on a Lafayette memory drum. The items were presented at a 2:2-sec. rate with a 4-sec. intertrial interval. Five different orders of presentation were utilized to minimize serial learning effects. All syllables used as responses were spelled; all words used as responses were pronounced. The *S*s did not verbalize the stimulus items.

RESULTS

The results are summarized in Table 1. An analysis of variance on the mean number of trials required to learn each of the familiarization lists revealed that the groups did not differ significantly, $F\,(3, 50) = .99$, $p > .05$. Except where indicated otherwise, two-tailed tests were used throughout the analyses and the assumption of homogeneity of variance was met. On the basis of this analysis it may be said that the familiarization lists for the Relevant and Irrelevant groups were of comparable difficulty level. A *t* test on mean number of trials required to learn the test list was performed between Control Groups 3 and 4 to determine whether there were significant differences in the difficulty of learning the criterion lists independent of familiarity training. This analysis indicated that the two lists did not differ in difficulty level, $t\,(13) = .60$, $p > .05$. A similar analysis on mean errors to criterion yielded comparable results, $t\,(13) = .26$, $p > .05$. In all in-

TABLE 1

MEANS AND *SD*s FOR ALL MEASURES

Measure		Group				
		RM	RNM	IRM	IRNM	C-3
First-list familiarization trials	*M*	22.00	20.00	22.47	19.80	—
	SD	6.20	4.62	6.79	5.17	—
Test-list trials to criterion	*M*	14.00	12.20	11.07	12.80	19.87
	SD	2.13	4.41	5.54	3.68	5.92
Total errors on test list	*M*	20.33	29.87	26.47	23.60	51.47
	SD	6.53	9.51	10.67	7.55	15.25
Correct anticipations on Test Trials 1–7	*M*	38.67	29.33	33.40	34.07	21.19
	SD	5.32	6.36	5.82	5.85	5.01

stances, an error was defined as either an incorrect response or a failure to respond. On the basis of these latter analyses the criterion data for Control Groups 3 and 4 were combined in all subsequent comparisons between these groups and the familiarization groups. In subsequent discussion the combined data of these control groups will be referred to simply as C-3.

To ascertain the effects of stimulus familiarization of test-list performance an analysis of variance was conducted on mean number of trials to reach criterion on the PA test list. In this analysis one *S* in Group IRM and two *S*s in Group C-3 failed to reach criterion and were assigned a score of 30. Results indicated a highly significant *F* ratio, $F\ (4,\ 70) = 80.87$, $p < .001$. Hartley's test revealed that the assumption of homogeneity of variance could not be met in this analysis, $F_{\max}\ (5,\ 14) = 7.66$, $p < .01$, but numerous studies have found that this violation does not seriously affect the application of this statistic (Winer, 1962). Individual *t* tests indicated that Groups RM, RNM, IRM, and IRNM were significantly superior to C-3 ($p < .001$) but did not differ significantly from each other. A similar analysis of mean errors to criterion again yielded highly significant differences between the groups, $F\ (4,\ 70) = 21.3$, $p < .001$. Again, the assumption of homogeneity of variance could not be met, $F_{\max}\ (5,\ 14) = 5.3$, $p < .05$. The results of *t* tests indicated that Group RM made significantly fewer errors than did Group RNM, $t\ (28) = 3.23$, $p < .01$, and that Group IRNM was just short of being significantly superior to Group RNM, $t\ (28) = 2.01$, $.10 > p > .05$. Groups RM, IRM, and IRNM did not differ significantly from one another. All of the stimulus familiarization groups made significantly fewer errors than did C-3 ($p < .001$).

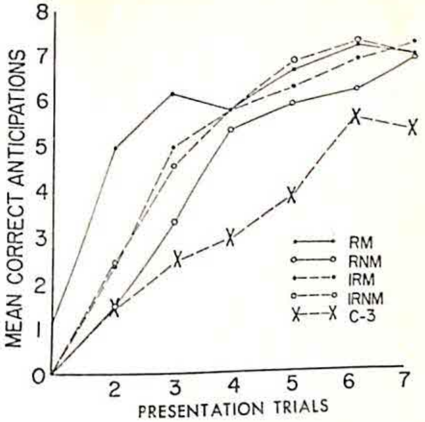


FIG. 1. Criterion paired-associate learning over Presentation Trials 1–7.

Figure 1 shows the mean number of correct anticipations for the mediated and nonmediated groups over Criterion Presentation Trials 1–7. This figure presents the results of the first seven trials only because the familiarization groups were performing at essentially the same level beyond the seventh trial. The fact that Group RM starts at 1.00 on the first trial reflects the finding that two *S*s in this group correctly guessed some of the dominant PA responses prior to the first anticipation trial. A repeated-measurements analysis of variance on the first seven trials for the mediated and nonmediated groups (excluding C-3) indicated that the Treatment, Trials, and Treatment × Trials interaction were all significant sources of variance ($p < .001$ for all three ratios). One *S* each in Groups RNM, IRNM, and IRM reached criterion in six trials and was assigned a score of 8 on the seventh trial. The significance level for the Trials effect was not altered by the Greenhouse and Geisser approximation procedure for the assumption of equal covariance (Winer, 1962). Orthogonal comparisons for the Trials and Treatment × Trials effects indicated significant linear components for Trials ($p < .001$) and Treatment × Trials ($p < .001$) and a significant quadratic component for Trials ($p < .001$). The quadratic component for the Treatment × Trials interaction was not significant.

Individual group comparisons on the mean number of correct anticipations for Trials 1–7 (see Table 1) revealed that Group RM made more correct responses than did Group RNM, $t\ (28) = 4.39, p < .001$; Group IRM, $t\ (28) = 2.60,\ p < .05$; and Group IRNM, $t\ (28) = 2.27, p < .05$. Groups IRM and IRNM did not differ significantly from one another. A Duncan's range test was performed to determine the significance of the means at each trial over the four treatment conditions. Trials 2 and 3 were the only trials on which significant differences were obtained. Group RM made significantly more correct anticipations on Trial 2 than did all of the other groups ($p < .01$) and was superior only to Group RNM on Trial 3 ($p < .05$).

DISCUSSION

The results of this study indicate that prior PA stimulus familiarization may either facilitate or hinder subsequent PA learning. While supporting the Underwood and Schulz (1960, Exp. 4) hypothesis, this latter finding goes considerably beyond their original formulation in terms of specifying more precisely the nature of the conditions under which PA stimulus familiarization may or may not inhibit criterion PA learning. Thus, in the present study, when the familiarized syllable reappeared as the stimulus in the test-list PA task *and* when the response to be learned in the test phase was not a dominant pre-experimental word associate of the stimulus employed in the familiarization phase (Group RNM), inhibition of test-list learning occurred. When the test-list syllable differed from the syllable familiarized on the pretest (Group IRNM), however, a general "learning to-learn" or "warm-up" effect was observed (see Group C-3).

As Murdock (1958) has suggested, the inhibitory effects obtained in Group RNM also may be interpreted within the general framework of negative transfer theory. Murdock hypothesized that if backward learning occurred during the learning of an A-B PA, then the learning of a second PA B-C should be hindered in accordance with transfer of training theory which predicts negative transfer in a B-A, B-C paradigm. Glaze 100% association-value nonsense syllables were used as both stimuli and responses in Murdock's study and his findings, in agreement with the results obtained in Group RNM, confirmed this latter inhibitory prediction. It should be particularly noted, however, that in the present study an A-B, B-C paradigm also produced facilitation effects in Group RM. This latter result is in general agreement with Storms' (1958) finding that backward associations inferred from word-association norms can reliably act as mediators in facilitating PA test-list learning. It is also consonant with the results of McCormack's (1961) studies which indicate that the backward associations of experimentally produced mediators can facilitate criterion PA tasks. Together with the inhibitory effects found in Group RNM, the facilitating effects obtained in Group RM indicate that the associative relationships existing among the stimulus and response materials employed in first- and second-list PA learning under an A-B, B-C paradigm are critical variables in determining whether prior PA stimulus familiarization will facilitate or inhibit criterion PA learning.

REFERENCES

CIEUTAT, V. J. Differential familiarity with stimulus and response in paired-associate learning. *Percept. mot. Skills*, 1960, 11, 269–275.

GANNON, D. R., & NOBLE, C. E. Familiarization (n) as a stimulus factor in paired-associate verbal learning. *J. exp. Psychol.*, 1961, **62**, 14–23.

McCormack, P. D. Backward mediated positive transfer in a paired-associate task. *J. exp. Psychol.*, 1961, **61**, 138–141.

Murdock, B. B., Jr. "Backward" associations in transfer and learning. *J. exp. Psychol.*, 1958, **55**, 111–114.

Noble, C. E. Measurement of association value (*a*), rated associations (*a*′), and scaled meaningfulness (*m*′) for the 2100 CVC combinations of the English alphabet. *Psychol. Rep.*, 1961, **8**, 487–521.

Russell, W. A., & Jenkins, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8-onr-66216.

Storms, L. H. Apparent backward association: A situational effect. *J. exp. Psychol.*, 1958, **55**, 390–395.

Umemoto, T., & Hilgard, E. R. Paired-associate learning as a function of similarity: Common stimulus and response items within the list. *J. exp. Psychol.*, 1961, **62**, 97–104.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received May 7, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 316–323

INTERPAIR INTERFERENCE AS A FUNCTION OF LEVEL OF PRACTICE IN PAIRED-ASSOCIATE LEARNING[1]

SAM C. BROWN[2]

University of Virginia

3 experiments, each designed to test the hypothesis that interference from other pairs in a paired-associate (PA) list first increases and then decreases during learning, involved the presentation of pairs to be learned in a list also containing either previously learned (L), unlearned (U), or new (N) pairs. While the results of Experiments I and II were inconclusive, Experiment III revealed maximal interference produced by U pairs of a medium degree of practice. Such interference was significantly and nonmonotonically related to the degree of previous practice on the interfering pairs, being minimal for N and L pairs. These results, in addition to demonstrating both the facilitory and inhibitory effects of repetition, argue against all-or-none interpretations of PA learning.

Interference as an explanatory concept in verbal learning has been commonly invoked in studies of retention and transfer (e.g., Postman, 1961, 1963a; Underwood, 1957; Underwood & Postman, 1960). However, relatively little attention has been directed toward interference as a phenomenon occurring within the learning of a single list, attributable to interactive effects among the various items constituting the list (Irion, 1959). A notable exception has been Gibson's (1940) formulation based on the concept of stimulus generalization, but this has been subjected to increasing criticism due to its restrictiveness and limited empirical relevance (Underwood, 1961), as well as the operational ambiguity of the concept of stimulus generalization itself (Battig, 1959; Gibson, 1959; Murdock, 1958, 1959; Runquist, 1959). The present research was directed toward a more generally satisfactory empirical description of the role of interference in paired-associate learning produced by other pairs within a single list, termed *interpair interference* (IPI).

A number of considerations, similar to those leading to Gibson's (1940) formulation concerning stimulus generalization, suggest that IPI should vary in magnitude over the course of learning, reaching a maximum at some intermediate stage of the learning process. That is, in order for items to interfere with the learning of one another, something must initially be learned about them. Likewise, insofar as *S* achieves errorless performance, IPI should decrease late in learning. Therefore, it seems reasonable to expect less IPI from pairs characterized by relatively low (e.g., new pairs presented for the first time) and high levels of performance (e.g., pairs previously responded to correctly) than from some class of previously presented pairs which have not been responded to correctly, designated herein as unlearned pairs. An experimental test of these predictions was the major concern of the present series of investigations.

---

[1] Experiment I, supported in part by the Cooperative Research Program of the United States Office of Education, is based in part upon a master's thesis, and Exp. II and III, supported both by a Public Health Service fellowship (MPM-18,527) and Grant M-5769 from the National Institute of Mental Health, Public Health Service, represent portions of a PhD dissertation, both submitted to the Graduate School of the University of Virginia in partial fulfillment of the respective degree requirements. The author is indebted to William F. Battig for his invaluable assistance during all stages of this investigation.

[2] Now at Kansas State University.

EXPERIMENT I

Method

Subjects.—The 80 paid *S*s were mostly summer session students at the University of Virginia, the majority of whom had previously participated in other verbal-learning experiments.

Part I procedure.—All *S*s practiced upon an identical paired-associate list of 12 nonsense-syllable pairs to a criterion of at least 6 out of 12 pairs correct. Each learning trial began with the successive presentation of all 12 pairs through an automatic 2 × 2 in. slide projector, followed immediately by a recall series, during which only the stimulus syllables of each pair were exposed in a different order, and *S* attempted to pronounce the appropriate response. During the recall series a shutter operated by *E* was lowered over the projection screen so that only the stimulus member of each pair was exposed. A 5-sec. rate (4-sec. exposure, 1 sec. between successive slides) was maintained throughout each trial, with a 30-sec. interval between trials and between alternating presentation and recall series within each trial, to permit rearrangement of the slide order for the next series.

Upon entering the experimental room each *S* was seated in front of a projection screen, which shielded him from *E* and the apparatus. Typical instructions for paired-associate learning were given, but no mention was made of the performance criterion.

Part II procedure.—After reaching criterion, each *S* was allowed a 5-min. rest interval during which time assignment was made to one of four groups on the basis that the number of *S*s (20), total errors, and variability of scores be matched for all groups. However, in order to more satisfactorily match the groups it was necessary to discard data of four additional *S*s, which was done solely on the basis of performance on Part I.

Following the 5-min. interval all *S*s practiced upon a second list of 12 paired associates, with the four groups differing in kind of previous experience with these pairs. Specifically, three sets of pairs were distinguished, designated as (*a*) learned (L) pairs (those previously responded to correctly during Part I); (*b*) unlearned (U) pairs (those previously seen during Part I but never responded to correctly); and (*c*) new (N) pairs which were never seen on Part I.

Each of the four groups learned a list representing a different combination of these L, U, and N pairs, as follows:

1. Group L-U (learned-unlearned). The *S*s continued to learn both the set of six L and six U pairs of Part I.
2. Group L-N (learned-new). The six L pairs of Part I were retained in the second list along with six additional N pairs replacing the U pairs.
3. Group U-N (unlearned-new). The second list consisted of the six U pairs from Part I along with six N pairs which replaced the set of L pairs.
4. Group N-N (new-new). The *S*s were required to learn 12 N pairs, 6 of which were the same as those used in Groups L-N and U-N.

Since it was not always possible to terminate *S*'s performance on Part I when exactly six pairs were responded to correctly, some of the pairs included in the U sets of Groups L-U and U-N were, in fact, pairs responded to correctly on Part I. In such instances, the sixth pair recalled correctly was treated as an L pair, and all remaining pairs designated as U pairs on Part II. The number of these pairs included in the U sets was approximately the same in both groups, being 1.1 and 0.9 per *S* in Groups L-U and U-N, respectively. However, an adjusted U score, employed for all Part II analyses, was tabulated separately for each *S* in Groups L-U and U-N. This consisted of the average number of errors per pair for those pairs in the U set which were not responded to correctly on Part I, multiplied by six, the number of pairs included in the U set on Part II.

The procedure used in Part II was identical to Part I, except that *S*s were required to learn the second list to a criterion of two successive perfect trials or to a maximum of seven trials. In addition, prior to presentation of the Part II list, all *S*s were told that some of the pairs in Part II might be new.

Materials.—Twenty-four nonoverlapping pairs of nonsense syllables, each with an average 50% association value (Archer, 1960), were selected from a larger pool of pairs used in a previous study (Brown & Battig, 1962, Exp. II). The 24 pairs were divided into four sets of 6 pairs which were approximately equal in mean number of errors made by different *S*s in the previous study, variability of these scores, and similarity of the items included within each set.

Two of the sets were combined into a single list of 12 pairs used in Part I, while the remaining two sets constituted the N pairs which were used in Part II. Moreover, one of these new sets (employed in Groups L-N and U-N) could be combined with any 6 pairs of the Part I list, or with the other set of N pairs, in a manner which would yield lists comparable in intralist similarity with the initial 12 pairs of Part I.

Results

The effects of IPI were determined by comparing the mean number of errors made on Part II on the same set when learned with different sets of pairs. While performance on the L set was slightly better in Group L-N (4.3) than in Group L-U (6.0), this difference failed to reach significance by analysis of covariance ($F < 1$) at the .05 level, the criterion of significance for all analyses to be reported. Likewise, little difference in performance on N pairs was evident in the U-N (16.0) and N-N (16.9) groups ($F < 1$). Therefore, the postulation that IPI would increase to a maximum early in learning was not supported. However, Group L-U (12.9) was significantly superior to Group U-N (17.9) on U pairs, $F(1, 37) = 7.72$, as was Group L-N (13.5) to Group N-N (16.9) on N pairs, $F(1, 37) = 4.64$. Similarly, fewer errors were made on the N set by Group L-N (13.5) than by Group U-N (16.0). Although this latter difference was not significant, $F(1, 37) = 2.34$, significantly fewer errors were made on the first three trials on the N set by Group L-N (10.1) than Group U-N (11.9), $F(1, 37) = 4.53$. Therefore, all comparisons consistently show L pairs to produce less interference than U and N pairs, in support of the original hypothesis.

Analyses of errors preceding (before errors) and proportion of errors following the first correct response to each pair (proportion of after errors), as well as mean errors on the first trial, all revealed results substantially the same as those obtained using total errors, except that only U pairs evidenced significant and differential sensitivity to interference on the first trial as determined by the L-U (3.4) and U-N (4.3) group comparison, $F(1, 37) = 4.78$.

In addition to the evaluation of IPI, performance on sets differing in previous practice was compared directly, when each was learned conjointly with the same set of pairs. Mean errors on U and N pairs proved to be virtually identical when compared between (*a*) Groups L-U (12.9) and L-N (13.5); (*b*) Groups U-N (17.9) and N-N (17.7 for both N sets combined); and (*c*) U (17.9) and N (16.0) sets within Group U-N. However, in each instance, U pairs displayed significantly fewer errors on the first two trials, and more on subsequent trials, than did N pairs, as shown by extended-trend analysis of variance (Grant, 1956), which revealed the slopes of the N learning curves to be significantly steeper than those for the U sets, all Fs $(1, 38) \geq 4.94$.

EXPERIMENTS II AND III

The failure to find maximal IPI produced by U sets in Exp. I may indicate that (*a*) the degree of learning of these U pairs had proceeded beyond the point of maximal IPI, particularly since some of the U pairs employed on Part II were actually L pairs, (*b*) maximal IPI is produced by N pairs after a single presentation, and such interference does not decrease substantially until relatively late in learning, and/or (*c*) the L and N pairs, upon which such IPI was measured, were insufficiently sensitive as indexes for detecting differential interfering effects. Consistent with the latter suggestion is the fact that only U pairs were found to be differentially interfered with on the first trial of Part II, at which point any differences in IPI between sets should have been most evident.

Consequently, modifications were introduced in Exp. II and III in order to (*a*) more adequately represent various degrees of learning by distinguishing between three different sets of U pairs varying in degree of previous practice; and (*b*) provide more sensitive indexes for the measurement of IPI, through the use of various classes of less difficult word-bigram pairs. Previously learned pairs were included for the attempted measurement of IPI in Exp. II, but were eliminated in Exp. III because of their demonstrated insensitivity for this purpose.

Method

Although two separate experiments were conducted, both procedures were basically alike and can most efficiently be described together.

Subjects.—The 120 paid *S*s, 60 in each experiment, were selected from the same population as that used in Exp. I. Three *S*s were eliminated for either failure to follow instructions or refusal to try to learn the list, and data from five other *S*s were discarded to more satisfactorily match the groups.

Part I procedure.—As in Exp. I, all *S*s practiced upon an identical list of pairs, although all pairs responded to correctly for the first time (L pairs) were discarded from the list at the completion of that trial on which that correct response occurred. In Exp. II, two additional pairs, randomly selected from those never responded to correctly (U pairs), were also removed from the list of each *S* along with the first two L pairs, and set aside for subsequent use on Part II. In Exp. III, three U pairs were removed, but with the first L pair. As a consequence, number of pairs remaining in the list underwent a progressive decrease from trial to trial, and total number of eliminated L pairs also served to define the various learning criteria upon which experimental manipulations in Part II were based (as described below).

Other procedural details were identical to those of Exp. I, except for a 3-sec. presentation and recall rate, and instructions to *S* to always pronounce the word and spell the two letters of the bigram of each pair. The *S*s were also told about the removal of the pairs from the list, although no mention was made of the particular contingencies involved.

Based upon either the number of correct responses on the first trial or the number of trials required to respond correctly to at least one pair, *S*s in both experiments were divided into five matched groups. The groups differed principally with respect to criterion of first-list learning, expressed in terms of total number of different pairs responded to correctly. In accordance with the degree of practice of the particular pairs of Part I used subsequently in the second list on Part II, the five groups in both experiments represent and will be referred to as (*a*) Group L (learned pairs), (*b*) Group U-H (unlearned pairs of a high degree of practice), (*c*) Group U-M (unlearned pairs of a medium degree of practice), (*d*) Group U-L (unlearned pairs of a low degree of practice), and (*e*) Group N (new pairs never seen during Part I). Table 1 presents the learning criteria, differing slightly for Exp. II and III, for each group. One third of the *S*s in Group N learned to each of the three criteria used for the corresponding unlearned groups.

Part II procedure.—After reaching criterion on the first list, each *S* engaged in a 5-min. number-cancellation task, during which time *E* rearranged the materials for use with the Part II list. The second list consisted of two sets of six word-bigram pairs each, designated, respectively, as IPI *measuring* pairs, which were equivalent for all groups and upon which IPI was actually measured, and IPI *producing* pairs, which were systematically different for each of the five groups and constituted the source of IPI. For Group L the producing

TABLE 1

PART I LEARNING CRITERIA (TOTAL PAIRS RESPONDED TO CORRECTLY) FOR EACH GROUP IN EXP. II AND III

Exp.	Groups			
	L	U-H	U-M	U-L
II	8 or more	8–10	5–7	2–4
III	7 or more	7–9	4–6	1–3

pairs were the last six L pairs from the first list of each *S* (all having been responded to correctly once on Part I). For the three unlearned groups, these producing pairs were randomly selected from the pool of unlearned pairs remaining in each *S*'s first list after reaching the appropriate first-list learning criterion. Since Groups U-L, U-M, and U-H learned the first lists to progressively increasing levels of mastery, thereby resulting in concomitant increases in the amount of practice on the remaining U pairs, these will be referred to, respectively, as unlearned-low, unlearned-medium, and unlearned-high producing pairs. For Group N, the same six new producing pairs were used for all *S*s in both experiments, none of which were seen on Part I.

In Exp. II, three subsets of measuring pairs were employed, two of which consisted of the first two L pairs and the two U pairs simultaneously removed from the first list, while the third was comprised of two N pairs which were identical for all *S*s, but selected so as to be easier to learn than other pairs in the list. Since preliminary analysis of Exp. II revealed virtually perfect performance by all groups on the two L measuring pairs, this subset was eliminated in Exp. III, and an additional pair was added to both the U and N subsets, thereby reducing the number of subsets to two. Consequently, in Exp. III, these subsets consisted of the three U pairs removed from the first list of each *S* at the time of the first correct response, and of three N pairs, two of which were identical to those used in Exp. II.

In each experiment, the procedure employed in Part II was identical to Part I, except that (*a*) learning of the second list was carried out to a criterion of one errorless trial or to a maximum of 18 trials; (*b*) presentation of stimulus terms for attempted response recall was under self-paced conditions, and (*c*) no pairs were removed from the list.

Materials.—Twenty-seven different word-bigram pairs, consisting of three-letter words of 98–100% association value (Archer, 1960) and two-letter bigrams of minimal frequency of occurrence in the English language, were selected from those used in a previous study (Battig & Nelson, 1962).

Eighteen of the pairs constituted the first (Part I) list for all *S*s, whereas 6 others represented the N producing pairs learned in Part II by *S*s in Group N. Both groups of pairs were approximately equal in mean proportion of correct responses and inter-*S* variability in the Battig and Nelson (1962) study. The remaining 3 pairs, constituting the N measuring pairs, were also homogeneous in difficulty of learning, but were, on the average, learned more rapidly by *S*s in the previous study. Moreover, these N measuring pairs could be combined with the 6 N producing pairs and/or any of the 18 pairs from the first list so as to yield lists of minimal and comparable intralist similarity on Part II.

Results

Some *S*s in both Exp. II and III exceeded the first-list learning criterion of the group to which originally assigned, as determined by the matching criterion. However, supplementary Part II analyses based either upon (*a*) reassignment of *S*s on the basis of actual first-list learning criteria, or (*b*) average rather than individual performance measures (for those *S*s assigned to different groups but who achieved the same first-list criterion) revealed results substantially the same as those based upon the original groups. Therefore, only data and analyses based upon the latter procedure will be reported.

Mean total errors made by the five groups in each experiment on the set of Part II measuring pairs are presented in Fig. 1. A consistent decrease in errors from new to learned producing pairs, with the three unlearned groups intermediate between

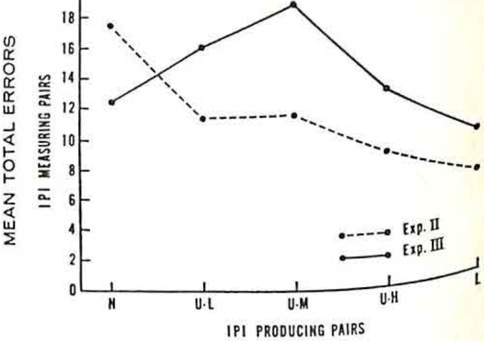


FIG. 1. Mean total errors on the set of II measuring pairs for each group of Exp. and III.

them, is clearly shown in Exp. II, leading to a significant linear component by extended-trend analysis of variance, F (1, 54) = 9.95, for the IPI function in Fig. 1. In contrast, errors on measuring pairs in Exp. III increased progressively from new through unlearned-low producing pairs, reaching a maximum for unlearned pairs of a medium degree of practice and decreasing again from unlearned-high to learned producing pairs. Only the quadratic component of the IPI function for Exp. III was significant, F (1, 54) = 4.59, thereby directly supporting the original hypothesis. All other orthogonal components of both IPI functions were insignificant (all Fs < 1). Measures of before and after errors, and of trials to criterion, were basically similar to those presented in Fig. 1.

That the relationship found in Exp. II was due primarily to performance on the U and N subsets of measuring pairs was indicated by the nearly identical shapes of the IPI functions for these two subsets, along with the virtually perfect performance of all groups on the two L measuring pairs. The shapes of the IPI functions obtained with the U and N subsets employed in Exp. III were also similar to one another, and no significant interactions involving the U and N curves were observed, all Fs (1, 55) $\lesssim 1.46$.

In agreement with similar comparisons made in Exp. I concerning the effects on performance of differences in previous practice, total errors made in Exp. III on the N (15.2) and average U (17.7) sets of producing pairs did not differ significantly by analysis of covariance ($F < 1$). However, U producing pairs displayed significantly fewer errors on the first trial of Part II and more on subsequent trials, than did N producing pairs, as indicated by a significant difference in the slopes of the two curves, F (1, 44) = 6.17. In Exp. II, however, not only were significantly more total errors made on the N (24.5) than on the average U (15.6) sets of producing pairs, F (1, 43) = 5.69, but this difference was maintained throughout Part II.

Additional comparisons of interest in both experiments showed (*a*) the IPI functions depicted in Fig. 1 to reflect primarily proportion of omission rather than overt errors; and (*b*) a consistent decrement in performance on unlearned-low through unlearned pairs of a high degree of practice (with the learned producing pairs markedly superior to all other sets). This latter finding suggests the differential IPI effects observed on measuring pairs not to be exclusively attributable to the greater difficulty of the U producing pairs, since more errors were made on unlearned-high than unlearned-low and medium producing pairs in both experiments.

The major discrepancy between the two experiments lies in the performance of Group N. This group in Exp. II made more errors on the six N producing pairs than did Group N in Exp. III on these same pairs, despite the lesser difficulty of the Exp. II list due to its inclusion of two L measuring pairs. Moreover, two *S*s accounted for 47% of the Group N errors on measuring pairs in Exp. II (although elimination of the data for these two deviant *S*s still failed to bring these differences in line with Exp. III). Consequently, the discrepancies between the two IPI functions must at least partially reflect the use in Exp. III of lower degrees of practice on producing pairs for the three unlearned groups, and of more sensitive measuring pairs.

DISCUSSION

Despite the differences between the three experiments, it may be concluded that the present results offer rather conclusive evidence for the existence of IPI in paired-associate learning, and clearly show that performance on any given pair is not independent of other pairs in the list. Furthermore, in view of the greater procedural adequacy and sensitivity of Exp. III, it also appears that such IPI is maximal for pairs of low to moderate degrees of practice and learning. To the extent that the latter findings are reliable, these results not only demonstrate both the facilitory and inhibitory effects of repetition, otherwise completely confounded in the typical paired-associate task, but also support the notion that the use of only one value of U pairs in Exp. I may have been responsible for the failure in that experiment to find differences of IPI between U and N pairs.

That the differences in IPI observed in the present experiments may reflect differential attention by *S*s to the various sets of producing pairs seems a definite possibility, particularly since in both Exp. II and III differential IPI was found to be primarily attributable to omission errors, which under the present self-paced conditions would reflect primarily inattention to these measuring pairs. However, it is not at all clear why *S*s would concentrate more on unlearned-low and medium producing pairs than on unlearned pairs of a high degree of previous practice. Nevertheless, the fact that minimal IPI was produced by learned pairs, in all three experiments, strongly suggests that *S*s may effectively ignore previously correct pairs and concentrate primarily on the remaining pairs in the list.

The fact that U pairs produced differential interference and that fewer errors were made on U than N pairs on early trials of Part II in Exp. I and III (despite the greater difficulty of these U pairs as evidenced by subsequent inferior performance on this set) clearly indicates that something had been learned about the U pairs on Part I, contrary to all-or-none interpretations (e.g., Bower, 1962; Estes, 1960; Rock, 1957), but in agreement with other investigations (e.g., Battig, 1962; Postman, 1963b; Underwood & Keppel, 1962; Williams, 1962). Moreover, since N pairs produced less interference than any class of U pairs employed in Exp. III, failure to find a facilitating effect of repetition of incorrect pairs under the Rock (1957) repetition condition may reflect the greater interfering effects produced by these U pairs as compared with the N pairs used in the nonrepetition condition, as also suggested independently both by Kristofferson (1961) and by Clark, Lansford, and Dallenbach (1960).

Although limited to a description of changes in IPI over the course of learning a single list of paired associates, the approach taken herein appears potentially useful not only for the investigation and precise specification of mechanisms responsible for IPI, but also for the elucidation of variables related to such IPI.

REFERENCES

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

BATTIG, W. F. Comment on "Intralist generalization in paired-associate learning." *Psychol. Rev.*, 1959, **66**, 338–339.

BATTIG, W. F. Paired-associate learning under simultaneous repetition and nonrepetition conditions. *J. exp. Psychol.*, 1962, **64**, 87–93.

BATTIG, W. F., & NELSON, D. Effect of kind of material and previous experience on paired-associate learning under repetition and nonrepetition conditions. *Canad. J. Psychol.*, 1962, **16**, 106–111.

BOWER, G. H. An association model for response and training variables in paired-associate learning. *Psychol. Rev.*, 1962, **69**, 34–53.

BROWN, S. C., & BATTIG, W. F. Partial serial position constancy in nonserial paired-associate learning. *J. verbal Learn. verbal Behav.*, 1962, **1**, 42–47.

CLARK, L. L., LANSFORD, T. G., & DALLENBACH, K. M. Repetition and associative learning. *Amer. J. Psychol.*, 1960, **73,** 22–40.

Estes, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67,** 207–223.

Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47,** 196–229.

Gibson, E. J. A re-examination of generalization. *Psychol. Rev.*, 1959, **66,** 340–342.

Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53,** 141–154.

Irion, A. L. Rote learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. II. New York: McGraw-Hill, 1959. Pp. 538–560.

Kristofferson, A. B. Repetition and paired associates learning. *Science*, 1961, **134,** 2036–2037.

Murdock, B. B., Jr. Intralist generalization in paired-associate learning. *Psychol. Rev.*, 1958, **65,** 306–314.

Murdock, B. B., Jr. A reply to Battig, Gibson, and Runquist. *Psychol. Rev.*, 1959, **66,** 345–346.

Postman, L. The present status of interference theory. In C. H. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961. Pp. 152–179.

Postman, L. Does interference theory predict too much forgetting? *J. verbal Learn. verbal Behav.*, 1963, **2,** 40–48. (a)

Postman, L. One-trial learning. In C. H. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963. Pp. 295–321. (b)

Rock, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70,** 186–193.

Runquist, W. N. Remarks on "Intralist generalization in paired-associate learning." *Psychol. Rev.*, 1959, **66,** 343–344.

Underwood, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64,** 49–60.

Underwood, B. J. An evaluation of the Gibson theory of verbal learning. In C. H. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961. Pp. 197–217.

Underwood, B. J., & Keppel, G. One-trial learning? *J. verbal Learn. verbal Behav.*, 1962, **1,** 1–13.

Underwood, B. J., & Postman, L. Extraexperimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67,** 73–95.

Williams, J. P. A test of the all-or-none hypothesis for verbal learning. *J. exp. Psychol.*, 1962, **64,** 158–165.

(Received May 18, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 324–329

PREDICTION OF AUDITORY DISCRIMINATION LEARNING AND TRANSPOSITION FROM CHILDREN'S AUDITORY ORDERING ABILITY[1]

DONALD A. RILEY, JOHN P. McKEE, AND RAYMOND W. HADLEY

University of California, Berkeley

Initially, 1st- and 3rd-grade children were pretested with visual stimuli and found to understand the concept of order. The Ss in each grade were then divided into 4 groups. Experimental Ss received in succession training in auditory ordering, then auditory discrimination learning, and finally a test of auditory transposition. For ½ the experimental Ss the stimuli were of different frequency, and for the other ½, of different intensity. Control Ss were treated the same as experimentals except that they were not given training in auditory ordering. 3rd-grade Ss ordered auditory stimuli more readily than 1st graders, and they also learned faster and verbalized more correctly. Intensity Ss ordered more readily than frequency Ss and also learned faster, transposed more, and verbalized more correctly. These experiments confirm the hypothesis that auditory discrimination learning and transposition are related to ability to place the stimuli in order.

Previous papers have reported that first-, second-, and third-grade children and college students learned a 12-db. loudness discrimination rather easily and generally transposed when tested immediately after learning (McKee & Riley, 1962; Riley & McKee, 1963). Age differences in learning and transposition were not significant. The results for similar *S*s trained in a 500-cycle frequency discrimination were different: the frequency problem was more difficult to learn; it became easier as age increased; and the frequency *S*s did not transpose significantly more often than chance. Indeed, when testing conditions permitted, frequency *S*s showed a significant tendency to make the absolute choice rather than the relative one when tested for transposition. Since the frequency difference was slightly more discriminable than the 12-db. amplitude difference, the learning differences could not be attributed to differences in discriminability.

The present experiments seek to determine whether the different results for frequency and amplitude are related to differences in *S*s' abilities to place such stimuli in the correct order from high to low or from loud to soft.

METHOD

Subjects

One hundred and thirteen first-grade children and 85 third-grade children from three different schools in Berkeley, California served as *S*s. All schools serve middle-class neighborhoods. Two of the three schools are the same as those used in the previous experiments. The *S*s, taken in alphabetical order, were assigned in rotation to the various experimental conditions. If an *S* failed to learn, the next *S* was assigned to that condition. Among first graders, 26 of the 66 *S*s assigned to frequency-discrimination groups failed to reach criterion and 7 of the 47 *S*s who had to make a discrimination based on amplitude failed. Among third graders 4 of the 44

---

[1] This research was supported in part by Grant 06811 from the United States Public Health Service and in part by a grant from the National Science Foundation to the Center for Human Learning.

We are greatly indebted to C. H. Wennerberg, Robert Muscio, Frank Wylde, and Orville Sipe of the Berkeley Public Schools for their helpfulness in making available children and testing facilities.

frequency *Ss* and 1 of the 41 amplitude *Ss* failed to learn.

Experimental Procedure

Visual pretraining.—To make certain that he could arrange stimuli in order when asked, each *S* was given visual pretraining in which he was presented with four strips of black cardboard .5 in. wide and 1, 1.5, 2.25, and 3.4 in. long and was asked to "Put these in the right order." If the child failed to respond he was asked to "Make them go the right way." If he still failed to respond *E* demonstrated, then mixed them up and said, "Now you do it." In addition, all *Ss* were given identical pretraining with four grays all of the same size. Their reflectances were 30, 24, 17, and 9%. Half the *Ss* began with the grays and half with the lengths. As soon as *S* was successful he was asked why he had arranged the stimuli as he had and then proceeded to the second half of the pretraining.

The four sizes were ordered correctly by 95% of the third-grade *Ss* and 82% of the first-grade *Ss* on the first trial. Ordering by brightness was slightly more difficult, particularly for first-grade *Ss*, but even so, all *Ss* were successful within five trials. For first-grade *Ss*, the mean number of trials required to reach one correct ordering was 1.8 and for third graders it was 1.2.

Auditory ordering and training.—Following visual pretraining *Ss* were divided into four groups. One group, Group F-C, which was a control for the effect of visual ordering, was trained on the same 500-cycle frequency discrimination used in previous studies; the second control group, Group A-C, was trained on the same 12-db. amplitude discrimination used previously. The third and fourth groups received auditory ordering prior to discrimination training. They were shown four identical $4 \times 3 \times 1\frac{1}{2}$ in. boxes containing $2\frac{3}{4}$-in. Galvin speakers that produced tones when they were lifted from the table. Hewlett-Packard and Grason-Stadler audio oscillators generated the tones. For the third group (Group F-E) the speakers emitted different frequencies—1,125, 1,500, 2,000, and 2,666 cps—all equated in loudness to the 2,000-cps tone which was approximately 85 db. at the speaker. For the fourth group (A-E) the speakers emitted a 2,000-cycle tone of approximately 109, 97, 85, and 73 db. at the speaker. (The amplitude drop from speaker to ear was about 20 db.) The apparatus was demonstrated to the child and auditory ordering then proceeded in exactly the same fashion as the visual pretraining. Auditory ordering continued for a maximum of seven trials or until the child responded correctly on two consecutive trials.

Following this auditory ordering, Group F-E learned a 500-cycle frequency discrimination and was then tested for frequency transposition. Group A-E learned a 12-db. amplitude discrimination and was then tested for amplitude transposition. Table 1 shows the overall design of these experiments which was applied to both first- and third-grade *Ss*. Twenty *Ss* were trained to criterion and tested for transposition in each of the four conditions in both grades.

In the discrimination learning part of the experiment, *Ss* were trained to a criterion of nine correct responses in 10 consecutive trials. Furthermore, every *S* was trained for at least 20 trials. If *S* was not at criterion performance on the twentieth trial, training continued until it was reached. The *Ss* who did not reach criterion in 40 trials were discarded.

TABLE 1

EXPERIMENTAL DESIGN

Procedure	Groups			
	F-C (Frequency-Control)	A-C (Amplitude-Control)	F-E (Frequency-Experimental)	A-E (Amplitude-Experimental)
Visual pretraining	Length and brightness	Length and brightness	Length and brightness	Length and brightness
Auditory ordering	None	None	Frequency	Amplitude
Auditory discrimination	Frequency	Amplitude	Frequency	Amplitude
Transposition test	Frequency	Amplitude	Frequency	Amplitude

In both experiments the same counterbalanced order of presentation (high or low tone first; loud or soft tone first) was used. In each experiment the high (or loud) tone was "right" for half the *S*s, and the low (or soft) tone was "right" for the others.

The *E* first sounded the two tones in close succession (about a 1-sec. interval) and asked *S* if he heard the two tones. When the child responded positively, *E* then told *S* that one of the tones was right and one was wrong. The *E* then sounded the right and wrong tones. He never described the tones in terms of loudness or pitch; he simply referred to them as "right" or "wrong." After a number of illustrations with both temporal sequences, *E* said, "Sometimes the first tone is the right one and sometimes the second tone is the right one." The *E* illustrated each possibility. He then said, "Now I am going to make a number of these and each time I want you to tell me whether the right tone is first or second." Trial-by-trial training then began. Reinforcements consisted of "right," "yes," "good for you," and so on. When *S* made several consecutive wrong responses, *E* reinstructed him and gave examples of the right and wrong tones.

Transposition testing.—Immediately following training, and with no pause in the procedure and no comment, each *S* who reached the criterion was tested with a transposition series of 10 trials, all of which were reinforced. Half of each training group was tested with the higher (louder) of the training stimuli and a yet higher (louder) test stimulus; the other half was tested on the lower (softer) of the two training stimuli and a yet lower (softer) test stimulus.

TABLE 2

NUMBER OF *S*S WHO FIRST CORRECTLY ORDERED AUDITORY STIMULI ON SUCCESSIVE TRIALS

	Trial 1	Later Trial	Never
First grade			
Frequency	8	10	14
Amplitude	13	9	0
Third grade			
Frequency	14	4	4
Amplitude	20	1	1

Note.—This table includes both the 80 *S*s who later learned the discrimination task and the 17 who did not.

RESULTS

Ordering.—Table 2 shows the trial on which *S*s were first able to order the auditory stimuli correctly. Both first- and third-grade *S*s ordered amplitudes more readily than frequencies. For the first grade, χ^2 (2) $= 13.79$, $p < .01$; for third grade, χ^2 (1) $= 5.03$, $p < .05$. In addition, there is improvement from the first to the third grade in both dimensions. For frequency, χ^2 (2) $= 8.04$, $p < .02$; for amplitude, χ^2 (1) $= 6.34$, $p < .02$.

Learning.—The effect of auditory ordering on speed of learning was evaluated against two sets of control conditions: learning scores from previ-

TABLE 3

PROPORTION OF *S*S LEARNING IN EACH CONDITION AND MEAN TRIALS TO LEARN FOR THOSE LEARNING THE DISCRIMINATION

	Ordering Training Prior to Learning								
	None[a]			Visual Only			Visual and Auditory		
	p	*M*	*SD*	*p*	*M*	*SD*	*p*	*M*	*SD*
First grade									
Frequency	.62	23.20	4.50	.59	23.95	5.86	.62	25.55	6.14
Amplitude	.97	21.90	4.40	.80	20.70	1.61	.91	20.45	1.12
Third grade									
Frequency	.96	22.08	5.10	.86	20.75	1.97	.91	22.80	4.39
Amplitude	1.00	20.83	1.94	1.00	20.30	1.82	.95	20.70	1.95

[a] Data from previous experiments (McKee & Riley, 1962; Riley & McKee, 1963).

TABLE 4

Effect of Ordering on Transposition

	First Test Trial Proportion of *S*s Who Transposed			Mean of 10 Transposition Trials		
	No Order[a]	Visual	Visual and Auditory	No Order	Visual	Visual and Auditory
First grade						
Frequency	.50	.80	.70	6.50	7.00	6.05
Amplitude	.81	1.00	1.00	8.60	9.45	9.35
Third grade						
Frequency	.62	.50	.75	6.40	7.60	8.25
Amplitude	.92	.95	.75	9.60	9.10	9.20

[a] Data from previous experiments (McKee & Riley, 1962; Riley & McKee, 1963).

ous experiments in which no ordering occurred prior to learning, and conditions in which *S*s ordered the visual stimuli but not the auditory stimuli prior to learning. Table 3 shows the proportion of *S*s who learned and number of trials to criterion for those *S*s. The introduction of visual ordering and of visual ordering plus auditory ordering had no appreciable effect on ease of learning. So far as learning is concerned, the new experiments replicate previous observations that first graders learn an amplitude discrimination more readily than a frequency discrimination and that by the third grade this difference has largely disappeared.

The relationship between ability to order the auditory stimuli and ease of learning could be examined in detail only in the first-grade frequency *S*s who received training in auditory ordering. Only in this condition (Group F-E of the design) were there enough *S*s who required more than one trial to order, or who failed to reach the criterion of ordering. In this condition *S*s who ordered on the first attempt learned more easily—both in terms of the number who learned and in terms of trials to criterion for the learners—than those who ordered later, while *S*s who never reached the criterion of ordering were poorest in learning. The differences, however, are not statistically significant. Since the training procedure involved at least 20 trials for each *S* regardless of how well he performed at the start, an additional analysis of only the first 10 learning trials was undertaken with the expectation that it would be more sensitive to the differences. The mean correct in the first 10 learning trials for those *S*s who had correctly ordered the frequencies on the first ordering trial was 8.62. For all other *S*s the mean was 5.67; $t\ (30) = 3.14,\ p < .01$.

Transposition.—Table 4 shows, for *S*s who learned, the effect of previous ordering experience on transposition. Data from previous experiments in which no ordering training was given are included for comparison. As before, more transposition occurs in amplitude than in frequency at both grade levels. Ordering experience, whether only visual or visual plus auditory, appears to increase transposition on Trial 1 in both modalities for the first graders but not for the third graders. To test the significance of this effect, the combined two first-grade conditions in the present experiment were compared with the previous first-grade results. For frequency, χ^2 (1) = 4.32, $p < .05$, and for amplitude, χ^2 (1) = 5.94, $p < .02$. For third graders, the proportion of frequency *S*s who transpose on the first trial is greater with visual *and* auditory pretraining than with visual pretraining alone, while for the amplitude *S*s the reverse is true. This interaction was significant ($p < .02$) when tested with the following formula: $t = (p_1 - p_2) - (p_3 - p_4) / \sigma[(p_1 - p_2) - (p_3 - p_4)]$, and no ready explanation suggests itself. Finally, ordering has no significant effect on the mean number of transposition choices over 10 test trials.

The relationship between ability to order and degree of transposition can be evaluated only for first-grade frequency *S*s, for it is only in this group that the individual differences in ordering ability are sufficiently large to be analyzed. The mean number of transposition choices of the 10 *S*s who most readily ordered the frequency stimuli was 7.7 while the mean for the 10 *S*s who ordered only with difficulty or not at all was 4.4, t (18) = 1.93, $p < .10$.

DISCUSSION

The present experiments confirm the hypothesis that the superiority of amplitude *S*s to frequency *S*s in learning and transposition is related to differences in the degree to which *S*s can correctly order stimuli in the two dimensions. There remains the question of interpretation.

In both the present and preceding experiments, first-grade *S*s learned the amplitude discrimination quite readily but had considerable difficulty with the frequency discrimination. The present findings show that first graders can order amplitudes easily but have difficulty ordering frequencies. Furthermore, questioning at the end of the experimental procedures revealed that in the first grade 35 of the 40 amplitude *S*s were able to verbalize the correct solution but only 23 of the 40 frequency *S*s verbalized correctly. In the third grade, 35 of the 40 amplitude *S*s verbalized correctly and 30 of the 40 frequency *S*s verbalized correctly.

One interpretation of these differences is that during the first 6 or 7 yr. of life children are much more often explicitly instructed and reinforced with respect to amplitude—and to differences in amplitude—than with respect to frequency and frequency differences. This interpretation amounts to the assertion that the ordering, learning, and verbalization differences between frequency and amplitude *S*s reflect transfer of previous amplitude training. Possibly the verbalization differences indicate that the transfer is effected by verbal mediators (Kendler & Kendler, 1962; Kuenne, 1946). Presumably, the smaller differences found in third-grade *S*s reflect increasing transfer effects in frequency as a consequence of musical instruction and similar experiences.

The preceding hypothesis will also account for the fact that first graders show more transposition in amplitude than frequency, but it will not account for the fact that previous experiments have shown this transposition difference to be equally apparent in second- and third-grade children and in adults. The presence of this difference over so wide an age range suggests two possibilities. The first is, as Stevens (1957) suggests, that the human *S* responds to the two dimensions in different ways because the mechanisms that mediate the two are innately different. The other possibility is that the difference occurs because the particular frequency differences used are more discriminable than the particular amplitude differences. Indeed, third-grade *S*s *do* transpose when trained and tested with a 100-cycle difference. On the other hand, adult *S*s rate a 100-cycle difference as smaller than a 12-db. difference and yet do not transpose when trained and tested with a 100-cycle difference. Although the latter fact appears to refute the discriminability hypothesis, it is not certain that the rating procedure is the most appropriate way to compare intervals in different sensory dimensions nor that adult data can be generalized to children.

Still another possible explanation of the lesser transposition in frequency *S*s is in terms of *S* selection. That is, some 40% of the first-grade frequency *S*s failed to learn, and it might be argued that absolute responding goes along with rapid learning. To check this possibility, the transposition scores of the first-grade *S*s who were in the top and bottom halves of the learning distribution were compared. In all four first-grade conditions of the present experiment, and in the previous first-grade data as well, the fast learners transpose slightly *more* than the slow learners. Consequently the difference in transposition between frequency and amplitude *S*s cannot be attributed to selection.

One other matter requires comment. It is probable that the ordering procedures used in the present experiments—particularly the combination of visual and auditory ordering—increase frequency transposition in third-grade *S*s. Table 4 is not entirely consistent with respect to this possibility, but an analysis of variance of the frequency data of the present experiments shows that the mean number of transpositional choices is significantly, $F(1, 76) = 4.15$, $p < .05$, greater in third-grade *S*s than in first-grade *S*s. This effect was not present in the previous experiments which did not entail ordering. If the effect is genuine, it may be due to a certain instructional quality of the ordering procedures.

REFERENCES

KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, **69**, 1–16.

KUENNE, M. R. Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, **36**, 471–490.

McKEE, J. P., & RILEY, D. A. Auditory transposition in six-year-old children. *Child Develpm.*, 1962, **33**, 469–476.

RILEY, D. A., & McKEE, J. P. Pitch and loudness transposition in children and adults. *Child Develpm.*, 1963, **34**, 471–482.

STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

(Received May 21, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 330–334

STIMULUS ASPECTS RESPONSIBLE FOR THE RAPID ACQUISITION OF REVERSAL SHIFTS IN CONCEPT FORMATION[1]

MARTIN HARROW

Yale University School of Medicine

The stimulus factors responsible for the superiority of reversal over nonreversal shifts in concept formation were analyzed. 2 factors confounded in reversal shifts, within-dimension shifts and response reversals, were isolated for experimental investigation. Using card-sorting tasks, groups making only within-dimension shifts learned more rapidly than groups making both within-dimension shifts and response reversals. Both groups learned quicker than comparable groups making between-dimension shifts. The results indicate that within-dimension shifts are an important factor in the rapid acquisition of reversal shifts, and response reversals may impede learning. In this transfer-of-training situation, positive transfer occurs when the stimulus dimension on which discrimination is based is kept constant, but the specific stimuli used are changed.

A number of previous experiments in concept formation, using human *S*s, have compared the speed of learning reversal and nonreversal shifts. The empirical results from this type of experiment often have been interpreted in terms of theoretical issues concerning the nature of certain types of human learning and concept formation (Buss, 1956; Gormezano & Grant, 1958; Harrow & Buchwald, 1962; Kendler & D'Amato, 1955; Kendler & Kendler, 1959; Osgood, 1953; Reese, 1962). In order to relate the results in this area to various theoretical formulations with greater exactness, there is need of a more precise analysis concerning which stimulus factors, or combination of factors, are responsible for the relatively quicker learning of reversal shifts by most older human *S*s.

In the experiments using sorting tasks *S*s were required to learn two successive card- or block-sorting tasks. During second-task learning, the nonreversal *S*s were given a task which was based on some dimension of the stimuli which was irrelevant during first-task learning. As opposed to this, during second-task learning the reversal *S*s were given a task which *E*s usually describe as being (in some sense) either the "opposite" or the "reverse" of the first task. A more careful analysis of the actual experimental tasks given to the "reversal" *S*s, however, shows that there were at least three ways in which their tasks differed from those given to the nonreversal *S*s.

1. Within-dimension shift (WDS). The second sorting task was based on the same dimension of the stimuli as was the first one (e.g., if the color of the cards was used as a basis for correct sorting during first-task learning, the color again formed the basis for correct sorting on second-task learning).

2. Response reversal (RR). The *S*s

---

[1] This paper is based on a dissertation submitted to the Graduate School of Arts and Sciences of Indiana University in partial fulfillment of the requirements for the PhD degree in psychology. The author is greatly indebted to A. M. Buchwald, W. K. Estes, I. Gormezano, and H. G. Yamaguchi for their valuable advice and suggestions.

were required to literally reverse the responses previously associated with each particular sorting card (e.g., if sorting Task 1 required that Response Card No. 9 be placed in Sorting Category A and Response Card No. 8 in Category B, then Sorting Task 2 required that Response Card No. 9 be placed in Sorting Category B, and Response Card No. 8 in Category A).

3. Concept reversal (CR). In several experiments (Harrow & Buchwald, 1962; Kendler & D'Amato, 1955; Kendler & Mayzner, 1956) the response cards were sorted into categories which were labeled with stimulus cards. In these experiments correct sorting depended on a systematic relationship between the response cards and the stimulus cards. Second-task learning involved a reversal of the systematic or conceptual relationship that had been used as a basis for sorting during Task 1 (e.g., response cards with stars located on them have to be sorted according to the number of stars on them. During Task 1 response cards having N stars must be sorted with the stimulus cards having $N + 1$ stars and during Task 2 response cards with N stars must be sorted with stimulus cards having $N - 1$ stars.)

Thus, due to the confounding in previous experiments there were three possible factors all or some of which could have been responsible for the quicker learning of reversal shifts as compared to nonreversal shifts. These were: (*a*) WDSs, (*b*) RRs, and (*c*) CRs.

In the present experiment, groups of *S*s making WDSs and other groups learning tasks which required both WDSs and RRs were compared for speed of learning with each other and with comparable control groups learning tasks requiring between-dimension shifts (BDSs). This permitted the separation of the effects of RRs and of WDSs from other possible influential factors.

METHOD

Subjects.—The *S*s were 118 undergraduate students enrolled in introductory psychology courses at Indiana University. Twenty-two of them were eliminated for failure to learn the first task within 160 trials, and the data from the remaining 96 (48 males and 48 females) were used.

TABLE 1

EXPERIMENTAL DESIGN

Groups[a]	Type of Shift[b]	Task 1			Task 2		
		Deck of Cards Used[c]	Correct Responses		Deck of Cards Used[c]	Correct Responses	
			Category I	Category II		Category I	Category II
1M & 1F	WDS & RR	A	green	red	C	red	green
2M & 2F	WDS	B	yellow	blue	C	red	green
3M & 3F	BD (Group 1 Control)	A	cross	star	C	red	green
4M & 4F	BD (Group 2 Control)	B	cross	star	C	red	green
5M & 5F	WDS & RR	A	cross	star	B	star	cross
6M & 6F	WDS	C	circle	triangle	B	star	cross
7M & 7F	BD (Group 5 Control)	A	green	red	B	star	cross
8M & 8F	BD (Group 6 Control)	C	green	red	B	star	cross

[a] M = male *S*s; F = female *S*s.

[b] WDS = within-dimension shift; RR = response reversal; BD = between-dimension shift.

[c] All decks had eight cards with the following combinations: Deck A, three or four stars or crosses, colored green or red; Deck B, one or two stars or crosses, colored blue or yellow; Deck C, one or two circles or triangles, colored green or red.

Materials.—Three decks, each with eight response cards, were selected from the asymmetrical version of the Wisconsin Card Sorting Test. The decks of cards used for each group, as well as the experimental design, can be seen in Table 1. The experiment was carried out in a two-category card-sorting situation in which *S*s had to sort response cards with one of two identical blank stimulus cards which were attached to a cardholder. There was a pocket at the top of the cardholder into which a response card could be placed. Two types of tasks (shape concepts and color concepts) were employed in the experiment. Two sets of response cards were used for each group, one for each of the two stages of the experiment.

Design.—A 2 × 2 × 2 × 2 factorial design was used, with the 96 *S*s being divided into 16 groups of 6 *S*s (see Table 1). All experimental *S*s learned two successive card-sorting tasks, with the four factors present during second-task learning being:

1. WDSs (Groups 1, 2, 5, and 6) vs. BDSs (Groups 3, 4, 7, and 8).
2. Relevant stimuli for correct discrimination not having been present during Task 1 (Groups 2, 4, 6, and 8) vs. relevant stimuli for correct discrimination having been present during Task 1 (Groups 1, 3, 5, and 7). (It should be noted in connection with this factor that the crucial aspect which distinguished RRs—Groups 1 and 5—was that the relevant Task 1 stimuli were again present and relevant during Task 2, but the specific responses to these stimuli were reversed. In contrast, the crucial factor which distinguished the groups making only WDSs, without RRs—Groups 2 and 6—was that the stimuli relevant during Task 2 had not been present during Task 1.)
3. Learning a color concept (Groups 1, 2, 3, and 4) vs. learning a shape concept (Groups 5, 6, 7, and 8).
4. Male *S*s (Groups 1M, 2M, 3M, etc.) vs. female *S*s (Groups 1F, 2F, 3F, etc.).

Viewed from a different aspect, the eight groups of male *S*s and eight groups of female *S*s differed from each other in the following ways:

1. Groups 1 and 5 (WDSs and RRs). These groups made a WDS with the specific stimuli relevant during Task 2 having been present and relevant during Task 1.
2. Groups 2 and 6 (WDSs). These groups made a WDS with the specific stimuli relevant during Task 2 not having been present during Task 1.
3. Groups 3 and 7 (BDSs, controls for Groups 1 and 5). These groups made a BDS with the specific stimuli relevant during Task 2 having been present but irrelevant during Task 1.
4. Groups 4 and 8 (BDSs, controls for Groups 2 and 6). These groups made a BDS with the specific stimuli relevant during Task 2 not having been present during Task 1.

Procedure.—The *S*s were run individually. The standardized instructions read to *S* indicated that as each response card was shown to him he should point to the category he thought it belonged to and by *E* informing him whether he was right or wrong he would gradually find out where each response card really belonged.

The response cards were presented randomly and individually to each *S* by being placed in the slot which was at the top of the cardholder. The criterion of learning for both the first and second tasks was 16 successive correct responses. All *S*s who met the criterion for the first task within 160 trials were required to learn a second one.

The *S*s were not informed about the change in the pattern of reinforcement for the second concept. The *S*s who had not learned the second concept within 160 trials were arbitrarily assigned a score of 160 for second-task learning.

RESULTS AND DISCUSSION

Table 2 presents the results of the learning of the first and second tasks. These data represent the number of trials to learn the task involved, excluding the 16 criterion trials. To test for differences in speed of learning the first task, two separate 2 × 2 analyses of variances were computed, one for the four groups that learned a color concept for their second task (Groups 1–4) and one for the four groups that learned a number concept for their second task (Groups 5–8). The two factors present in both analyses were: (*a*) color vs. number concept and (*b*) deck of cards used. Neither of the analyses showed any significant main effects or interaction, indicating that the groups were equated with each other for practice effects prior to second-task learning.

To determine differences in speed of learning the second task, a 2 × 2 × 2

TABLE 2

NUMBER OF TRIALS TO LEARN TASK 1 AND TASK 2

Group	Type of Tasks[a]	Type of Shift[b]	N	Trials to Learn							
				Task 1				Task 2			
				M	SD	Mdn.	Range	M	SD	Mdn.	Range
1	C-C	WDS & RR	12	30.5	41.2	5.0	0–123	6.6	11.6	1.0	1–38
2	C-C	WDS	12	30.3	34.7	19.0	0–113	2.3	3.1	1.0	0–10
3	S-C	BD (Group 1 Control)	12	48.3	43.4	40.0	0–118	8.3	8.7	5.0	0–31
4	S-C	BD (Group 2 Control)	12	28.3	31.4	19.0	0–100	6.4	6.9	4.5	0–22
5	S-S	WDS & RR	12	34.4	33.7	20.5	1–87	8.9	10.0	6.5	1–32
6	S-S	WDS	12	22.6	23.3	18.5	0–80	4.4	9.0	1.5	0–32
7	C-S	BD (Group 5 Control)	12	59.8	58.0	45.0	0–141	31.3	44.2	19.0	3–160
8	C-S	BD (Group 6 Control)	12	26.0	36.2	8.5	2–127	38.4	48.9	18.5	2–160

[a] C = color task; S = shape task.
[b] WDS = within-dimension shift; RR = response reversal; BD = between-dimension shift.

× 2 analysis of variance was carried out. To obtain homogeneity of variance the data had to be subjected to a logarithmic $(1 + X)$ transformation (Edwards, 1950). The results of the analysis (Tables 2 and 3) show that the WDSs were learned significantly faster than the BDSs. Likewise, color concepts were learned significantly faster than shape concepts. The only significant interaction was that between the type of shift and color-shape concept. Examination of Table 2 shows that this interaction was significant because of the extremely slow learning of the BDS groups that acquired a shape concept during Task 2.

The superiority of all groups making WDSs over the comparable ones making BDSs agrees, in principle, with the results obtained by Harrow and Buchwald (1962). In the present experiment this occurred in both types of situations requiring WDSs (i.e., when WDSs occurred in conjunction with RRs and when RRs were not present).

To assess the role of RRs and to contrast them with WDSs, further analysis of the four groups making WDSs (Groups 1, 2, 5, and 6) was

TABLE 3

$2 \times 2 \times 2 \times 2$ ANALYSIS OF VARIANCE OF TRIALS TO CRITERION ON TASK 2

Source	df	MS	F
Within-dim.—between-dim. shift (S)	1	4.943	25.399***
Relevant stim.-irrelevant stim. (R)	1	0.649	3.335
Color-shape concept (C)	1	2.771	14.237***
Males-females (M)	1	0.025	
S × R	1	0.413	2.121
S × C	1	0.877	4.505*
C × M	1	0.350	1.801
R × C × M[a]	1	0.210	1.077
Within groups	80	0.195	
Total	95		

Note.—Transformed scores [log $(1 + X)$] used.
[a] The F ratios for all remaining simple and higher-order interactions were less than 1, and are omitted.
* $p < .05$.
*** $p < .001$.

TABLE 4

Analysis of Variance of Trials to Criterion for Groups Making Within-Dimension Shifts

Source	*df*	*MS*	*F*
Type of within-dimension shift (S)	1	1.049	6.213*
Color-shape task (C)	1	0.265	1.571
S × C	1	0.122	
Within groups	44	0.169	
Total	47		

Note.—Transformed scores [log $(1 + X)$] used.
* $p < .05$.

carried out. This was accomplished in a separate 2 × 2 analysis of variance, using the transformed scores (log $1 + X$). The two variables present (see Table 4) were: (*a*) type of WDS, and (*b*) color-shape task. The essential comparison was that for the type of WDS. This comparison was significant, indicating (Tables 2 and 4) that Groups 2 and 6, which only made WDSs, learned quicker than Groups 1 and 5, which made both WDSs and RRs.

The data suggest that what appears to be the facilitating effects of RRs results from their always involving WDSs. There are further indications that RRs actually retard learning as compared to WDSs.

The results from the present experiment also permit us to answer a question basic to the nature of WDSs and of importance in concept formation. That is, if WDS lead to facilitative effects in learning, is this because *S*s have been trained to respond to the specific stimuli they have learned are important? The alternative hypothesis is that the advantages of a WDS occur because *S*s have learned to pay selective attention to a *general* dimension of the stimuli (such as color) while learning a first task, and that they continue to pay selective attention to that aspect of the stimuli, even when the *specific* colors formerly present are no longer there. In the current transfer-of-training situation, facilitation occurred for the groups making WDSs when the dimension of the stimuli on which discrimination was based was kept constant, but the specific stimuli important for correct discrimination were changed. This supports the hypothesis that the facilitative effects of a reversal shift occur because it involves using a dimension of the stimuli which was previously relevant.

REFERENCES

Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.

Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

Gormezano, I., & Grant, D. A. Progressive ambiguity in the attainment of concepts on the Wisconsin Card Sorting Test. *J. exp. Psychol.*, 1958, **55**, 621–627.

Harrow, M., & Buchwald, A. M. Reversal and nonreversal shifts in concept formation using consistent and inconsistent responses. *J. exp. Psychol.*, 1962, **64**, 476–481.

Kendler, H. H., & D'Amato, M. J. A comparison of reversal and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

Kendler, H. H., & Mayzner, M. S. Reversal and nonreversal shifts in card-sorting tests with two or four sorting categories. *J. exp. Psychol.*, 1956, **51**, 244–248.

Kendler, T. S., & Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58**, 56–60.

Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

Reese, H. W. Verbal mediation as a function of age level. *Psychol. Bull.*, 1962, **59**, 502–509.

(Received May 23, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 335–341

AN INVESTIGATION OF VARIABLES IN JUDGMENTS OF RELATIVE AREA[1]

HARRY HELSON AND WILLIAM BEVAN

Kansas State University

In 6 different experiments the part-whole proportion illusion reported by Paterson and Tinker was confirmed. The absolute error was most pronounced with proportions of intermediate magnitude but the relative error was greatest with smallest proportions of central area to total area. The error was found to be essentially invariant with changes in the range of relative magnitudes, the orientation of the rectangular stimulus area, and overall field size. However, when a null method of measuring the illusion was employed, the error was severely reduced with an inflection point at a proportion of 70%. The importance of contextual factors in judgments of a relative area is indicated by an inverse relationship between judged size and overall field size of the stimulus series when the comparisons are made within Ss rather than across groups of Ss. This last finding has methodological importance in determining the role of contextual factors in judgment.

A literature of long standing indicates that in judgments of area the constant error is usually one of overestimation (Anastasi, 1936; Paterson & Tinker, 1938). Curious about the tendency of *S*s to overestimate the proportion of the printed page occupied by print (printers, as a rule of thumb, set up about 50% of a book page as marginal space), Paterson and Tinker (1938) found that *S*s overestimated rectangular areas consisting of white or black on cards of contrasting color on representative book-page areas, about 18%, and actual printed matter in a wide variety of books by 25%. These writers simply referred the overestimations to the "part-whole proportion illusion." A number of possible explanations or sources of this tendency of judgment immediately suggest themselves: according to Piaget's (1947) Centration Effect, centrally fixated objects appear larger than objects in the periphery; Hollingworth's (1910) Central-Tendency Effect suggests that if the range of areas were extended, larger areas should be underestimated; the vertical-horizontal illusion suggests that perhaps the tendency to overestimate the vertical compared with the horizontal may be responsible since printed pages and the cards used by Paterson and Tinker were rectangular; the illusion may be a function of total field size; and there is the possibility that it may be an artifact of the method of measurement or the type of judgment demanded of *S*s. Accordingly, six experiments are reported in this paper investigating each of these possible explanations of the part-whole proportion illusion.

EXPERIMENT I

The first series of experiments began with the speculation that this error of overestimation might have resulted from a limited range covered by the stimulus series. Paterson and Tinker

---

[1] These experiments were performed under Contract Nonr-3634(01) between Kansas State University and the Physiological Psychology Branch, Office of Naval Research. The writers are indebted to Joan Faye Pritchard, Edward D. Turner, and James R. Haines for their assistance during the course of the work.

had a five-card series with the relative size of the central area varying from 42.7% to 71.9%. It was conjectured that if the series had been extended to cover a wider range, the Central-Tendency Effect might have been in evidence. Hollingworth, in his first test of this principle (1910), employed the estimation of the size of squares. More recently, a demonstration of the effect involved rectangles (Bevan & Dukes, 1953).

Experiment I thus consisted simply of the construction of a psychophysical function for judged relative area with two extended ranges of stimuli.

Method

Subjects.—Forty students in introductory psychology served as *S*s. They were divided at random into two groups except that there were 10 men and 10 women in each.

Materials.—Group I judged a series of 15 cards. Each card was 6 × 8 in. overall. These dimensions represent the median width and length of 100 books measured at random in the Kansas State University library. The cards were cut from heavy white posterboard (Crescent No. 600). The centers to be judged were made of black matt construction paper. The black areas judged in these experiments differed from the black areas of printed pages in being solid and were like the stimuli used by Paterson and Tinker in this respect. The center areas ranged from 0.41% to 88.0% of the total area of each card. Group II judged a 21-card series with centers ranging from 0.02% to 96.4% of the total area.

Procedure.—Each card was presented 10 times for judgment. Group I thus received a total of 150 trials, with the order randomized in blocks of 15. Group II received 210 trials randomized in blocks of 21. (In all of the present experiments the stimulus series consisted of either 150 or 210 trials depending on whether the 15-item or the 21-item series was used.) Testing was done in groups. The method was that of absolute judgment, *S*s being instructed to record their estimates of the proportionate size in per cents of the black centers on a record sheet provided for this purpose. The stimulus cards were concealed from *S*'s view and presented, one every 10 sec., for judgment. For each presentation the stimulus card was placed at eye level on a card rack for 5 sec.

Results

Figure 1A presents the average estimated per cent black plotted against actual per cent black for the 15-card series. Figure 1B presents similar data for the 21-card series. It is immediately apparent that the overestimation described by Paterson and Tinker

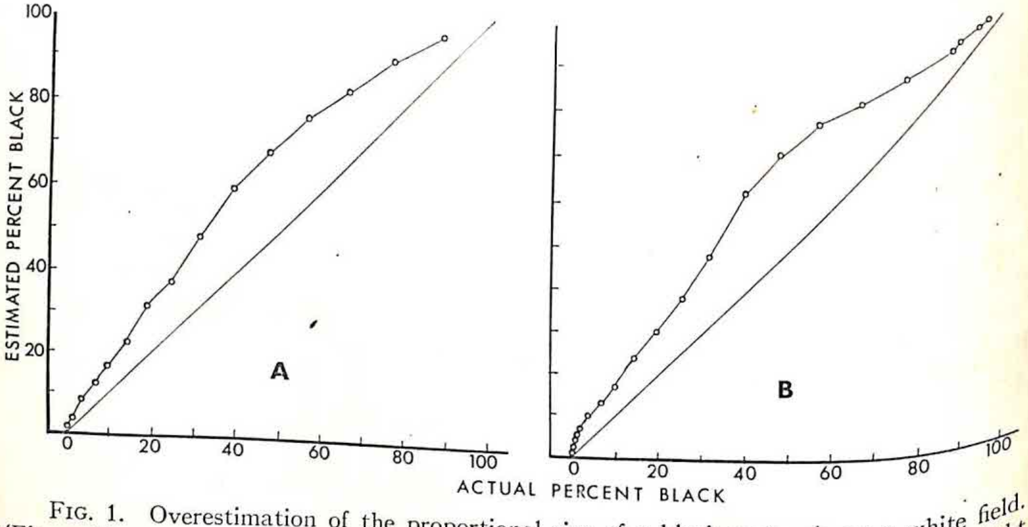


FIG. 1. Overestimation of the proportional size of a black rectangle on a white field. (Figure 1A presents data for the 15-item series Fig. 1B for the 21-item series. The straight line indicates veridical judgment.)

cannot be attributed to limited stimulus range, for overestimation predominates when almost the complete range of possible proportions is included. It is also clear that the greatest absolute overestimation occurs in the intermediate range of relative areas. For Group I the error of overestimation averaged 12.2%, with a range from 1.3% to 21.6%. In Group II the average was 9.5% with a range from 0.5% to 24.5%. Plots of the *relative* magnitude of the error (the ratio of the estimated to the actual per cent black) against actual relative size, result in a curve that is monotonic and negatively accelerated with the greatest effect at the low end and the smallest at the high end of the series. Thus in the 15-item series the log relative error is 2.49 for Card 1, and .92 for Card 15.

EXPERIMENT II

With a clear demonstration that overestimation is not attributable to limited series range, an intuitive search for variables related to this constant error was begun. One that immediately suggested itself was field size. The stimulus cards of Exp. I were the same size as an average-sized book, and it was conceivable that habitual attitudes associated with books could have been responsible for the obtained judgments. Paterson and Tinker had varied overall stimulus size, but only within a narrow range. Their series were 4 × 7 in., 5 × 8 in., and 6 × 9 in., respectively. Accordingly, a second experiment was undertaken.

Method

Subjects.—Sixty *S*s were involved in Exp. II, Group I of Exp. I and two additional groups of 10 men and 10 women each.

Procedure.—The 15-item series was reproduced at one-sixteenth and 16 times its original areas. One of the additional groups judged 1.5 × 2 in. cards, the other 24 × 32 in. cards. The procedure was that of Exp. I.

Results

Again, there was only overestimation. The average magnitude of error (13.3%) was approximately that of Exp. I, and there was no difference in the amount of overestimation as a function of field (card) size—1.5 × 2 in., 14.0%; 6 × 8 in., 12.2%; 24 × 32 in., 13.6%; F_{groups} (2, 54) = 1.01, $p > .05$. There was, however, a tendency for the female *S*s to display slightly larger error than the males—14.9% vs. 11.7%; F_{sexes} (1, 54) = 8.02, $p < .01$—that is reminiscent of the greater field dependence reported for females by Witkin, Lewis, Hertzman, Machover, Bretnall, and Wapner (1954).

EXPERIMENT III

Another plausible consideration in accounting for the results is the role of the horizontal-vertical illusion. If one dimension of a plane figure is overestimated, one might expect the relative area associated with this figure to be overestimated, since area is a product of linear dimensions. Earlier Peters (1933) reported that overestimation increased with an increase in the disparity of linear dimensions in rectangles, and Anastasi (1936) found a correlation of .83 between degree of overestimation and the greatest linear dimension. This latter, however, could have reflected a simple relationship between judged area and actual area. Accordingly, the role of the horizontal-vertical illusion was checked in a third experiment by having figures judged with the long dimension in both the vertical and horizontal orientations.

Method

Subjects.—Forty *S*s were included in Exp. II, Group II of Exp. I, and an additional group of 10 men and 10 women.

Procedure.—Data obtained with the rectangles of the 21-item series presented in the vertical orientation (Group II) were compared with judgments of the same series turned 90°. The procedure was that of Exp. I.

Results

Overestimation of relative error averaged slightly greater for the vertical orientation than for the horizontal (12.0% to 9.8%). However, this difference clearly failed to be significant, $F_{\text{groups}}\ (1, 36) = .02$; and thus the horizontal-vertical illusion cannot, with confidence, be regarded as a determinant of this constant error in areal judgment. Furthermore, there were no systematic sex differences in judgment, $F_{\text{sexes}}\ (1, 36) = 2.93$, $p > .05$, in these observations.

EXPERIMENT IV

The conditions considered to this point as determinants of the error of overestimation have been stimulus related. Response-related processes also suggest themselves as possible sources of error. Piaget (1947) insists that all perceptual—at least, visual—judgments have one essential property in common, the systematic overestimation of fixated items. This he calls the Centration Effect. Thus perceptual space is conceived to be spatially expanded at the focus of attention and progressively more contracted toward the periphery of the perceptual field. The fact of attention shifts causes multiple centration effects with compensation for the field distortion associated with a single fixation. Therefore, if the overestimation of the black centers was due to processes underlying centration, then a comparable effect should be produced for the white surround by having *S* fixate upon this portion of the visual field.

Method

Subjects.—Forty *Ss* were used. Group II of Exp. I served as the control group for center estimation. A second group of 10 men and 10 women was used for border estimation.

Procedure.—The control group had been instructed to fixate upon the central area of each card and estimate its relative size. The experimental group was now told to look at the white borders of the card and indicate the proportion of white to the total stimulus area.

Results

The centration hypothesis was not confirmed. While the control condition displayed the usual overestimation (12.0%), the group judging the white border showed a complementary error of underestimation of comparable magnitude (10.4%; $F_{\text{groups}} < 1.00$). Again, there was no significant sex difference in judgment ($F_{\text{sexes}} < 1.00$).

EXPERIMENT V

Experiment V was concerned with the error of overestimation as a reflection of method. The procedure utilized heretofore involved at least two judgmental operations: a comparison of areas in the stimulus field and the expression of this comparison in a conceptual system of per cents. Thus it is conceivable that the constant error is less a matter of perception than one of judgment of what constitutes specific relative magnitudes of particular dimensions. In a recent applied psychophysical experiment, Trumbo, Adams, Milner, and Schipper (1962) had *Ss* make simple stimulus judgments, with estimates in per cents, of various properties of systematically programed heterogeneous samples of wheat. They found a general tendency to overestimate the proportions of undamaged wheat of different grades in each sample, and, at the same time, they found constant errors of underestimation of the proportion of foreign material and damaged grain present. Therefore, Exp. V involved a null method, in which *Ss* matched the areas to be judged against a standard series.

Method

Subjects.—Twenty-one undergraduates, 11 women and 10 men, served as *Ss*.

Procedure.—The 21-item series of Exp. I was used as stimulus material and served also as standards. The *Ss* were shown these cards, one at a time, 10 times each, randomized in blocks of 21 trials, for a total of 210 trials. After each 5-sec. presentation, *S* identified one card from among the 21 cards in the standard series judged to be the same as the experimental card. The cards of the standard series were arranged in random order on a wall at a 90° angle to the left of the position of the experimental card and were presented at eye level, mounted against a soft tan background. Since classical time-order error involves comparison of a set of variables with *one* standard, the conditions used here minimized, if they did not completely obviate, this type of error.

Results

Figure 2 presents the results of Exp. V. Several things are immediately apparent. The change in psychophysical method is associated with a marked reduction in the magnitude of the error. Maximum overestimation, however, is the same with both the null and absolute methods and occurs at 40% area. In addition there is an inflection point in the function in the region of 70% relative area with larger black centers being slightly underestimated. However, the error is still predominately positive, F (1, 20) $= 6.29$, $p < .01$, and the inflection point is curiously displaced above the center of the series. Meanwhile, if this function is taken at face value it is suggestive of a central-tendency effect, with values below the inflection point being typically overestimated and those above underestimated. Again, there was no evidence of a sex difference in judgment ($F < 1.00$).

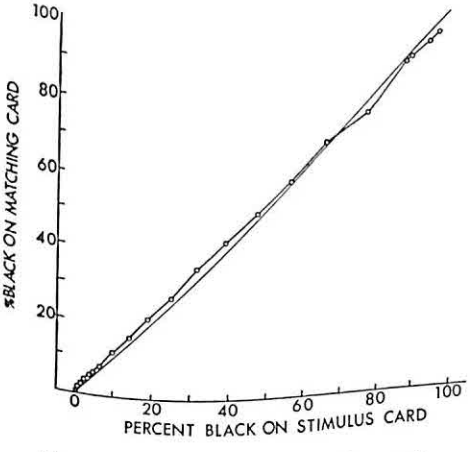


FIG. 2. Judgments of relative area involving the null method.

EXPERIMENT VI

Experiment II indicated that field size had no effect upon the magnitude of the overestimation. However, the adaptation-level literature (Helson, 1947) has demonstrated that the role of contextual factors is most readily identified when these variables are allowed to vary within the experience of the individual. Experiment I followed the common procedure of varying context across groups while holding it constant within groups. Experiment VI was performed as a complement to Exp. II. Here the relative size of the central black area was held constant and the overall card size was varied over the series.

Method

Subjects.—Sixty undergraduates, 30 men and 30 women, were used as *Ss*. They were divided at random into three equal-size groups.

Procedure.—Each group made a total of 70 judgments, randomized in blocks of seven. The overall dimensions of the seven series stimuli varied from $1\frac{1}{2} \times 2$ in. to $10\frac{1}{2} \times 14$ in., with each next larger card being $1\frac{1}{2}$ in. wider and 2 in. taller than the preceding card. For one group the central black area represented

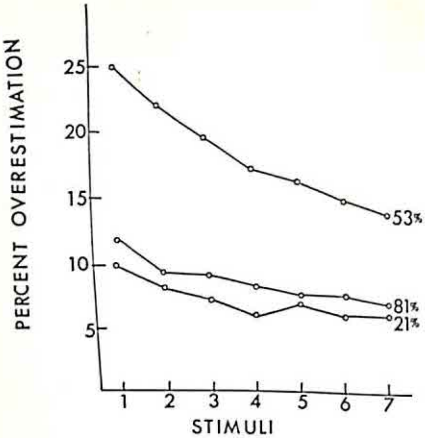


FIG. 3. Error of overestimation with judged area held physically constant and size of background varied. (Background size was increased from Stimulus 1 to Stimulus 7. The parameters indicate relative size of the area judged.)

21% of the total area of the card; for the second, 53% and for the third, 81%. The procedure was the same as that used in Exp. I and II.

Results

As in the earlier experiments, the error of overestimation was pronounced (Fig. 3). Again, it was greatest for the intermediate-sized black area (53%) and about the same for the complementary high and low values (81% and 21%). Finally, unlike the results of Exp. II, where each *S* was allowed to experience no variation in overall stimulus size, the error of overestimation is here seen to be clearly related to stimulus field size in an inverse fashion, the largest error in the case of all groups being associated with the smallest overall field size (Fig. 3). These results were statistically confirmed, F_{stimuli} (6, 324) = 41.41, $p < .01$; $F_{\text{black areas}}$ (2, 54) = 14.46, $p < .01$; $F_{\text{S}\times\text{BA}}$ (12, 324) = 5.32, $p < .01$. They are reminiscent of the data of Bevan, Pritchard, and Reed (1962) on the direction of series effects with loud and soft series of tones.

DISCUSSION

The present series of experiments confirmed the overestimation of relative area reported by Paterson and Tinker (1938) by exploring a wider range of situational variables, both stimulus and response, in its production. It is manifestly clear from the results that this constant error is strong and pervasive over a variety of conditions. Furthermore, as might be expected, it is greatest for relative magnitudes of intermediate size.

Experiments I through V strongly suggest that response variables are more potent determinants of the error than are stimulus variables. While the error did not vary as experimental groups differed in overall range of relative areas judged, overall field size, and orientation of the stimulus figures, a shift in psychophysical method from single-stimulus to the matching method, with a coincident change in the response language from a statement of proportion to that of simple discrimination was accompanied by a marked reduction in the constant error and the appearance of an inflection point above the midpoint of the series. However, the latter at 70% is so markedly discrepant from both the middlemost series value (32%) and the average stimulus size (36%), that it emphasizes the strength of the error and casts doubt on the inflection point as an indicant of the Central-Tendency Effect.

On the other hand, the importance of contextual factors anticipated by AL theory is demonstrated in the final experiment. Here field size as an independent variable was allowed to vary within the experience of each *S* and the overestimation was then found to be inversely related to total field size. This suggests an important methodological rule of thumb: since psychophysical judgments occur within a judgmental frame of reference, their competent investigation calls for designs in which stimulus variables are manipulated *within* rather than *across* groups if the role of contextual factors is to be identified.

REFERENCES

Anastasi, A. The estimation of area. *J. gen. Psychol.*, 1936, **14**, 201–225.

Bevan, W., & Dukes, W. F. Color as a variable in the judgment of size. *Amer. J. Psychol.*, 1953, **66**, 283–288.

Bevan, W., Pritchard, J. F., & Reed, W. G. Single-stimulus judgments of loudness as a function of presentation interval. *Amer. J. Psychol.*, 1962, **75**, 612–618.

Helson, H. Adaptation level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

Hollingworth, H. L. The central tendency of judgment. *J. Phil.*, 1910, **7**, 461–469.

Paterson, D. G., & Tinker, M. A. The part-whole proportion illusion in printing. *J. appl. Psychol.*, 1938, **22**, 421–425.

Peters, W. Versuche über den Einfluss der Form auf die Wahrnehmung der Flächengrösse. *Z. Psychol.*, 1933, **129**, 323–337.

Piaget, J. *The psychology of intelligence.* London: Routledge & Kegan Paul, 1947.

Trumbo, D. A., Adams, C. A., Milner, M., & Schipper, L. Reliability and accuracy in the inspection of hard red winter wheat. *Cereal Sci. Today*, 1962, **7**, 62–71.

Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, K., Bretnall, P. M., & Wapner, S. *Personality through perception.* New York: Harper, 1954.

(Received May 23, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 342–351

DIFFERENTIAL PREDICTION AND POSTDICTION OF WIN-LOSE EVENTS IN A SPATIAL-GENERALIZATION PROBLEM[1]

STANFORD H. SIMON

Veterans Administration Hospital, Tomah, Wisconsin

"Horses," represented by a row of 5 stimulus boxes, raced separately. Ss predicted whether horses won or lost. The most frequent event (E_1) was 80% for all horses. The type of E_1 for the central horse differed from the rest. 20 Ss had win for the center E_1 and 20 Ss had lose. ½ of each group was differentially motivated by being told it was their horses that were racing. After 150 trials, Ss postdicted the win frequency for each horse. Results show (a) Ss significantly underpredicted lose events as compared with win events and did so for each horse, (b) complex effects of the motivational variable, (c) a significant experimentally produced position preference effecting both Prediction and Postdiction with a related high incidence of rational predictions.

Brown, Clark, and Stein (1958) introduced a new technique for studying spatial generalization with voluntary responses. In this procedure seven lights were used, each signaling the racing of a different horse. The *S*'s task was to predict if the horse would win or lose. The center horse won 80% of the time and the rest won 20%. The record of win responses showed a gradient generalizing out from the center. Dixon and Wickens (1961) in a similar study noted:

> It is rather surprising that the Horse Race should produce a gradient. . . . Since there are 6 peripheral lights where losing predominates . . . it might be expected that generalization would take the form of suppressing the frequency of win responses for the central light, but in both groups the guessing frequency is close to the objective probability, 83% and 77% . . . [while] "win" responses are given at greater than the objective frequency for the peripheral lamps. Apparently, *S*s would rather win by betting on a winner than a loser [p. 509].

These differences between win and lose responses in matching objective frequencies are not seen in the similar data of Evans (1961), Fager and Knopf (1958), nor in the data originally published by Brown et al. (1958), although they are suggested in the estimates of win percentages (Postdiction) for each horse given by the Brown et al. *S*s after the experiment. In any case, none of the studies included a control group with the center horse a loser and the others winners. Such a control group was included in the present study which had as its primary purpose the examination of the effects of win-lose biases on Prediction and Postdiction responses in a spatial-generalization problem.

Two of the studies mentioned above included anxiety as a motivational variable with Dixon and Wickens (1961) inducing anxiety, and Fager and Knopf (1958) using groups differing in indigenous anxiety. In general, neither study found a significant effect of anxiety on spatial generalization. It was the second purpose of the present study to examine a different

---

[1] This research project was supported in part by the Veterans Administration Medical Research Program (8,200 funds). Portions of this paper were read at the 1963 Midwestern Psychological Association meeting in Chicago.

kind of motivational variable. One group was assumed to be more highly motivated by being told that they were the owners of the horses that were racing. A second group, assumed to be less highly motivated, received instructions that made the horses' ownership anonymous.

While the above studies have not been described as such, they can be viewed as representing one of a class of problems that would come under the purview of concurrent multiple-probability learning, i.e., the learning of a number of probability problems at the same time. Except for a study by Brody (1958), the author knows of no other work in this area. In Brody's study, *S*s were asked to predict which of two reinforcing lights would follow a signal light. His *S*s made two guesses following each signal light. He had two groups. Both groups had the same probability for the E_1 (most frequent event) for their first guess, but the E_1 for their second guess differed from their first guess and from each other. He concluded that "there was independence in the learning of two consecutive responses per trial . . . the asymptotic level of learning depended only upon the probability of reinforcement [p. 20]." In contrast to Brody's procedure, the present procedure has the probability problems identified in terms of differential spatial locations.

To make this research more comparable to previous work in probability learning, each "horse" unit contained its own feedback signals for E_1 and E_2 events and had two response keys rather than using only one pair of feedback signals common to all and a single toggle switch as had been used in all previous spatial-generalization studies of this type (except for Evans, 1961, who used two response keys).

An E_1 analysis of the data, in contrast to an analysis in terms of the generalized response as has been used in all previous studies, would reduce the range of scores between horses from an expected (with perfect matching) 20%–80%, to just 80%. Also, it would allow the effect of trials to be included in the analysis. It was the third purpose of this study to compare the sensitivity of these two types of analyses.

One other variable was studied. During the analysis of the data, an experimentally produced position preference emerged. This variable will be referred to as the Preferred Side variable, from which we identified the two groups: Preferred Right and Preferred Left.

METHOD

Apparatus.—The apparatus used was a modification of that reported by Brown et al. (1958). The *S* sat at a table and faced a black wooden panel on which was mounted a row of five metal boxes each 6 × 4 × 5 in. high. The boxes were at eye level for a seated *S*. The panel was curved so as to make the boxes spaced at 8° intervals and equidistant (5 ft.) from *S*'s face. On each box were three lights which formed an inverted V: a white light was centered 1 in. from the top, and two green lights were located $2\frac{1}{2}$ in. from the top and $1\frac{1}{4}$ in. from each side. A 3 × 6 in. white card with the words WIN and LOSE covered the green lights, each of which illuminated a different word. The cards could be changed so that WIN (or LOSE) could appear on the left or on the right, but for each *S* the WIN sign always appeared on the same side for all five horses. On top of each box a 2-in. high gray plastic toy horse stood against a 4 × 3 in. white cardboard background. On the table in front of *S* were two black push-button response keys, each mounted on a black plastic base. A control apparatus located in another room allowed *E* to turn on any one of the five white lights. When *S* pushed either one of the response keys, the white light went off and was immediately followed by one of the two green lights on the same box, the green light illuminating the WIN or LOSE sign for 1.5 sec. For all horses the E_1 event was always $p = .80$

and the designation of the E_1 event for the center position was different from the E_1 event for the other four horses (e.g., if the center horse was a winner, the rest were losers).

Design.—The experiment was originally designed as a 2 × 2 factorial design. For the Outcome variable, half the *S*s had the center E_1 = WIN (LWL) and half had the center E_1 = LOSE (WLW). For the Motivation variable, half of these two groups had the horses' owner anonymous (OA) and the other half were told that they were the owners (O). The side on which the WIN (and LOSE) signs occurred was counterbalanced within groups. Since the side on which the most frequent, i.e., noncentral, E_1 occurred proved to be a very significant variable, the general design became a 2 × 2 × 2 factorial, with the Preferred-Side variable divided between Preferred Left (PL) and Preferred Right (PR). To avoid confusion and errors, the left key always predicted the left event and the right key the right event.

Procedure.—After being seated, *S* was given the following instructions with the phrases in parentheses being used for *S*s in the Ownership (O) group:

> This is an experiment to see how well people can predict in a pretend horse racing situation. Each one of these boxes represents a different (one of your) horse. The horses are all from the same (your) stable and they run (you run them) one at a time against other horses not shown here, but they never run (you never run them) against each other.
>
> Whenever one of these white lights goes on, it means that horse is running (you are running that horse) in a race and you are to predict whether that (your) horse will win or lose. If you think he will win, you are to press this key like this with this hand (*E* demonstrates). Try it. If you think he will lose, you are to press this key with this hand like this (*E* demonstrates). Try that. Immediately after you have made your guess, you will find out what that (your) horse did. If he won, then the win sign for that horse will light up and if he lost, then the lose sign for that horse will light up. Now at first, you will just have to guess, but if you keep track carefully of how each of the (your) horses is doing, your guesses should become better and better. Remember, the (your) horses do not race against each other, so what one horse does has nothing to do with what the other horses do. The only important thing is to keep track carefully of how each horse is doing so that you can improve your guesses.

Ten practice trials were given, during which each horse won and lost once and the 10 events were randomized differently from *S* to *S*. During these trials, *E* verbalized the outcomes, e.g., "That (your) horse won." Following these trials, the key points in the last paragraph of the instructions given above were reiterated. Then *S* was left alone in the room for the learning trials.

There were 150 trials occurring at the rate of about 10–12 per minute. These trials were divided into six blocks of 25 trials each. Each horse was presented in random order five times per block. The E_1 event for each horse, therefore, occurred four times in each block of trials. The *E* had 10 cards with a different randomized block of 25 trials listed on each. After each *S*, the cards were shuffled and then the top 6 cards were used for the next *S*. (A comparable procedure was used for the practice trials.)

After finishing his trials, each *S* was asked to fill out a questionnaire on which he estimated the percentage of win for each horse, circled the horse he thought he had paid most attention to, and wrote the reason for doing so.

Subjects.—There were 40 *S*s, including 14 females, who were assigned at random to each condition, with 1 or 2 females in each cell. All *S*s were office personnel ranging from clerks to division chiefs in the Personnel, Fiscal, Registrar, Housekeeping, and Supply Divisions of the Tomah Veterans Administration Hospital. Ages ranged from 25 to 52 yr. with a mean of 41.3 yr. Education ranged from 8 to 17½ yr., with a mean of 12.6 yr.

RESULTS

Two kinds of analyses were made: one, in line with all previous studies, was oriented toward spatial generalization (generalization analysis) and the other toward multiple-probability learning (E_1 analysis). In the former analysis, the data used were the win responses for Groups LWL and the lose responses for Groups WLW for all five horses with all responses, therefore, being the same as the E_1 event of the central horse. The total scores on the last five of the six blocks of trials were used in these analyses.

Three analyses were made: one on *S*s' responses predicting the win-lose events (Predict data) using the actual scores, one on the estimate of percentage win made at the end of the experiment (Postdict data), and one on the differences between the two. For the last two analyses, percentage scores were used.

The analysis made in terms of multiple-probability learning used the scores of the E_1 events for each horse. With the Predict data, the analysis included the learning over the six blocks of trials. Again, the analysis on the Postdict data and the Predict – Postdict differences used percentage scores of E_1 events.

It should be noted that while the Postdict data were obtained in terms of the percentage of win for each horse, they were transformed to percentage of lose for each horse for those groups where E_1 for the center was lose (i.e., WLW group) in the generalization analysis, and for each horse when that horse had E_1 = lose in the multiple-probability-learning analysis.

Table 1 presents the mean scores for each horse (position) expressed in percentages of E_1 responses for all eight subgroups for both the Predict data obtained during the experiment (on the last five blocks of trials) and the Postdict data given immediately after the experiment. In this table as in all figures, the horses are numbered 1 to 5 from left to right. As can be seen, there is a marked difference in matching the objective probability (80% for all positions) between Position 1 and Position 5 in favor of the one that corresponds to the side on

TABLE 1

MEAN RESPONSES IN PERCENTAGE OF E_1 EVENTS FOR PREDICT AND POSTDICT DATA

	Preferred Left					Preferred Right				
	Horses					Horses				
	1	2	3	4	5	1	2	3	4	5
					Predict[a]					
Group OA										
LWL	76.8	64.0	56.0	46.4	52.0	67.2	70.4	76.0	76.8	79.2
WLW	88.0	76.0	65.6	57.6	67.2	72.0	74.4	64.0	82.4	89.6
Group O										
LWL	77.6	76.0	72.8	65.6	62.4	53.6	60.0	62.4	55.2	60.0
WLW	89.6	86.4	47.2	59.2	51.2	72.0	71.2	48.0	77.6	83.2
					Postdict					
Group OA										
LWL	63.0	64.0	49.0	56.0	75.0	66.0	71.0	76.0	72.0	74.0
WLW	81.6	76.0	66.0	53.0	68.0	66.0	55.0	76.0	69.0	80.0
Group O										
LWL	87.4	74.6	71.8	54.0	51.0	72.4	66.8	53.8	69.6	73.0
WLW	80.0	71.0	65.0	47.0	61.0	60.0	52.0	65.0	58.0	79.0

Note.—Probability of E_1 = .80.

[a] Predict data based on last five of six blocks of trials with each horse represented five times for each block.

which E_1 of the four noncentral horses occurred. For example, if E_1 was LOSE for Horses 1, 2, 4, and 5, and was located over the left green light of every box (horse), then E_1 for Horse 1 would approximate .80 more closely than Horse 5. The generalization data can be obtained by subtracting the scores of the noncentral horses from 100.

The results of the two outcome groups (WLW and LWL) for both the Predict and Postdict responses are presented in Fig. 1 for the generalization data (based on the last five blocks of trials), and the E_1 data (based on six blocks of trials) in Fig. 2. In the generalization data, the scores for all positions are in terms of the E_1 event for Position 3, i.e., the generalization response. For the E_1 data, the scores are in terms of the E_1 event for each position.

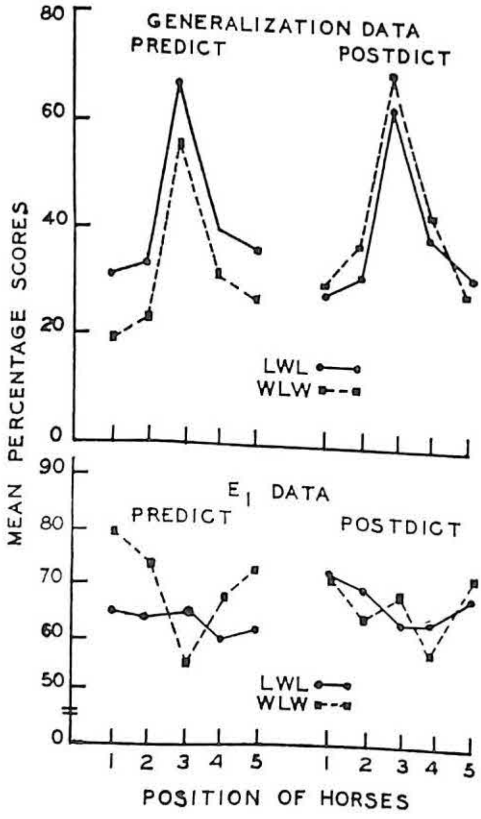


FIG. 1. E_1 and generalization curves of the two outcome groups for predict and postdict data.

To conserve space, only the E_1 analysis will be presented at length. Where appropriate, comparative statements will be made about the generalization analysis.

Analysis of E_1 responses.—Figure 2 presents the curves of the mean percentage of E_1 responses for the four major groups for both the Predict (six blocks of trials) and Postdict data. Table 2 presents an analysis of variance performed on the Predict, Postdict, and Predict-minus-Postdict E_1 data. Included in the table for comparative purpose is the significance level of all significant findings from the generalization analysis. These are denoted by the use of asterisks within parentheses. The actual figures of the generalization analysis are not presented. Except for the fact that the generalization analysis does not include the effect of trials, the two analyses run parallel with identical degrees of freedom.

The major findings are listed below.

1. The Outcome variable in the Predict data was significant ($p < .05$) with the WLW group getting closer to the 80% level than the LWL group. This is also evident in the Predict data interaction of Horses × Outcome which was significant at the .001 level. As can be seen in the E_1 data of Fig. 1, the two curves cross over with all E_1 = Win events having more E_1 responses than E_1 = Lose. This difference is significant at the .05 level with a Duncan range test for every position of the horses. These findings show that *S*s significantly underpredict lose events as compared with win events, and they do so for each position of the horses.

2. The findings in Paragraph 1 for the Predict data do not occur in the

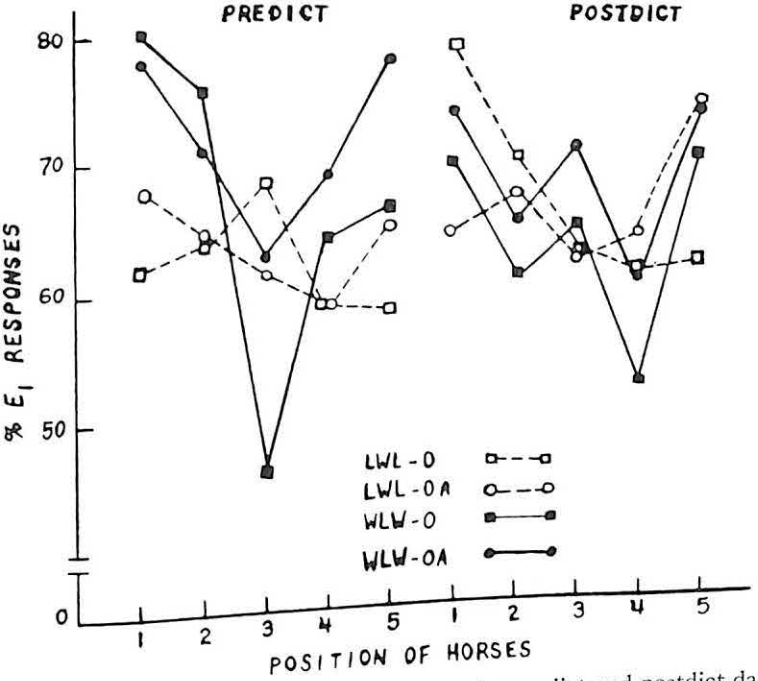


FIG. 2. E_1 responses for the four major groups for predict and postdict data.

Postdict data, but in the Predict-minus-Postdict data, Horses × Outcome is significant at the .001 level. The data show that when E_1 = Win, the Predict data are consistently, i.e., for every position, closer to .80 than the Postdict data—WLW Predict − Postdict is significantly larger than zero with $t\ (19) = 2.38$, $p < .05$; without the middle L horse $t = 3.45$, $p < .01$. When E_1 = Lose, the Postdict data are consistently closer to .80 than the Predict data, but the difference is not significant. (Since in Fig. 1 the Predict generalization curves do not cross over as do the E_1 curves these same effects are found in the Outcome variable with the generalization analysis.)

3. The effect of Horses, i.e., the different positions of horses, was significant at the .001 level for the Predict data and at the .01 level for the Postdict data. For the Predict data, the horses ranged from best to least in matching behavior in the following order: 1, 2, 5, 4, and 3, with the Duncan range test showing a difference significant at the .05 level between all pairs except 2–5. This shows the difference between the peripheral horses on each side to be significant and the order is congruent with the generalization effect. The generalization effect shows up to a lesser extent, i.e., only on one side, in the Postdict data where the Duncan range test shows only that Positions 5 and 1 are both larger than 4 at the .05 level.

4. The strong effect of the experimentally produced position preference is seen in the Horses × Preferred Side interaction which is significant at the .001 level for the Predict data and at the .01 level for the Postdict data. As can be seen in Table 1, for both the PL and PR, there are more E_1 responses for the two horses on the preferred side than for the two horses on the nonpreferred side.

5. The Trials effect is significant at

TABLE 2

ANALYSIS OF VARIANCE OF E_1 RESPONSES FOR PREDICT, POSTDICT, AND PREDICT-MINUS-POSTDICT DATA

Source	*df*	Predict		Postdict		Predict − Postdict	
		MS	*F*	*MS*	*F*	*MS*	*F*
Between							
Outcome (O)	1	30.1	$5.72^{*(***)}$	17.41		1,965.64	$1.73^{(**)}$
Motivation (M)	1	6.8	1.29	244.21		85.80	
Preferred Side (PS)	1	4.3		202.01		25.20	
O × M	1	2.7		465.12		117.05	
O × PS	1	1.6		411.84		1,109.21	
M × PS	1	25.8	4.90^{*}	554.44		812.05	
O × M × PS	1	23.6	4.49^{*}	239.81		920.20	
Error	32	5.26		768.25		1,135.53	
Within							
Horses (H)	4	14.38	$5.33^{***(***)}$	896.66	$3.79^{**(***)}$	773.38	3.42^{**}
H × O	4	13.98	5.18^{***}	275.54	1.17	1,707.09	7.56^{***}
H × M	4	2.60		285.47	1.21	268.27	1.19
H × PS	4	36.55	$13.54^{***(***)}$	1,562.99	$6.61^{**(*)}$	241.10	1.07
H × O × M	4	5.65	2.09	241.29	1.02	723.11	3.20^{**}
H × M × PS	4	.88		574.01	2.43^{*}	703.91	3.12^{**}
H × O × PS	4	1.80		240.28	1.02	200.44	
H × O × M × PS	4	1.30		252.97	1.07	514.12	2.28^{*}
Error 1 (w)	128	2.70		236.29		225.88	
Trials (T)	5	18.38	11.56^{***}				
T × O	5	.34					
T × M	5	.52					
T × PS	5	1.38					
T × O × M	5	.80					
T × O × PS	5	1.08					
T × M × PS	5	3.00	1.89				
T × O × M × PS	5	3.28	2.06				
T × H	20	1.66	1.34				
T × H × O	20	.88					
T × H × M	20	2.42	1.95^{*}				
T × H × PS	20	1.41	1.14				
T × H × O × M	20	1.71	1.38				
T × H × M × PS	20	1.00					
T × H × O × PS	20	1.41	1.14				
T × H × O × M × PS	20	1.79	1.44				
Error 2 (w)	160	1.59					
Error 3 (w)	640	1.24					

Note.—Asterisks in parentheses indicate significance level for analysis of generalization data. Figures for generalization data are not presented.

$^{*}p < .05$.

$^{**}p < .01$.

$^{***}p < .001$.

the .001 level indicating the learning that takes place.

6. A review of Table 2 shows that the E_1 analysis yields 16 significant differences, while the type of analysis based on generalization data used in all previous studies yields only 6. In only one instance did the latter analysis show a significant difference where none was found by the E_1 analysis. However, the relationship involved is to be found elsewhere in the E_1 analysis as was brought out in Paragraph 2 above. It may be noted that the primary points brought out in the first four major findings are also

to be found in the generalization analysis. This does not include the significant generalization effect in Paragraph 3. Also, the generalization analysis as such does not logically lend itself to a study of the effect of Trials, because of the discrepancy between the central horse and the others.

In addition to the above findings, there are a number of significant complex interactions, the meaning of which is not clear. These relationships with their significance levels are as follows: Predict data: Motivation × Preferred Side, Outcome × Motivation × Preferred Side, and Trials × Horses × Motivation, all with $p < .05$; Postdict data: Horses × Motivation × Preferred Side with $p < .05$; and Predict-minus-Postdict data: Horses, Horses × Outcome × Motivation, and Horses × Motivation × Preferred Side all with $p < .01$; and Horses × Outcome × Motivation × Preferred Side with $p < .05$. The relationships among the variables involved can be worked out from the data in Table 1.

DISCUSSION

The results show clearly that *Ss* will more accurately predict Win events than Lose events and they do this for each position of the "horses." Further, since this does not hold true for postdiction, and since the significant differences found between prediction and postdiction are based primarily on the underpostdiction of Win events, one cannot make a case that *Ss* "really knew" that the horses were losing more than they were predicting.

The motivational variable had no significant independent effect. However, in the E_1 analyses, it proved to be a significant factor in the following number of interactions: three for Predict data, one in the Postdict data, and three in the Difference data. All but one of these involved triple or quadruple interactions. There are no ready theories to account for these complex relationships.

The experimentally produced position preference proved to be a very significant variable. It is not clear what underlies this phenomenon. For descriptive purposes, it was identified as being related to the side on which the most frequent E_1 occurred (i.e., the side of the E_1 event for the noncentral horses). It may be that this variable is related to a perceptual phenomenon. For example, the sequence of events having the center signal light terminate and followed immediately by one of the side feedback lights tends to produce a phi phenomenon. This effect of movement occurring more often in one direction than the other for each *S* may have been the critical factor to lead him to pay more attention to the horses on the involved side. However, in order to avoid confusion and errors, the procedure required that the left hand always predicted the left event and the right hand the right event. The more frequent use of one hand as compared to the other may have produced a set that was related to *S* following the horses on one side more closely than the other. Evans (1961), however, presents data which seem to argue against this. He had left key = Lose and right key = Win. His data for four groups, all of whom were similar to our LWL, but all of whom had the apparatus partially obscured by a screen, showed the right side closer to matching than the left side. This difference is not commented on, and does not show up in a group without the screen. In any case, the relative importance of visual and motor variables on the experimentally produced position preference needs further clarification.

It is of interest to note that when *Ss* were asked after the experiment if they paid attention to any one of the horses more than any of the others, the end horse on the preferred side (PR or PL) was selected by 18 *Ss*, the nonpreferred side by 6 *Ss*, the middle horse by 8 *Ss*, and 8 *Ss* indicated no preference. (No *S* picked the second or fourth horse.) This yields a chi square of 7.7, significant at .05

level with $df = 2$. Further, the preferred side produced more rational behavior, i.e., 100% prediction of the E_1 event, than for any of the other positions. Table 3 shows the number of rational, matching, and other responses on the last block of five trials for each position from the preferred side to the nonpreferred side. As can be seen, the amount of rational behavior decreases as one goes from the preferred side to the nonpreferred side, and the amount of matching behavior, to a lesser extent, tends to do the same. A chi square of 28.39, significant at the .001 level with $df = 8$, is obtained from these data. Of particular interest is the large amount of rational behavior, especially for the end horse on the preferred side (60%; since Lose events were underpredicted, this jumps to 70% when only WLW groups are considered). It occurs on the last 5 of 30 trials. Certainly it is more than one would expect if only a single probability problem was being presented and with many more trials (see review by Estes, 1962, pp. 131–132).

It is tempting to relate the amount of rational predictions in this setting to factors involved in Siegel's (1959, 1961) theory of probability learning. He includes the two concepts: the utility of a correct response and the utility of variability, or the negative utility of "boredom." For any probability problem, as the utility of variability decreases, the asymptotic response probability increases. In the present situation, for example, the *S* who pays particular attention to the E_1 events for one of the horses (the one on the preferred side), may have the utility of variability in regard to the preferred horse reduced by the opportunity of varying his responses to the other horses, especially the center one, obtaining a maximization of E_1 responses for the horse on the preferred side. This may be especially true when two response keys are used instead of a single toggle switch. Evans (1961) also showed better-than-matching E_1 responses for the center as well as end horses in his data which does not show a position preference. His procedure also differed from previous studies, but was similar to the present one in his use of two response keys.

TABLE 3

TYPES OF PREDICTIONS ON LAST FIVE TRIALS FOR EACH POSITION RELATING TO PREFERRED SIDE

Position	Types of Prediction			Total
	Rational	Matching	Other	
PS	24	10	6	40
PS-1[a]	15	16	9	40
Center	13	8	19	40
NPS-1[a]	10	7	23	40
NPS	10	10	20	40
Total	72	51	77	200

Note.—PS = preferred side; NPS = nonpreferred side.

[a] Positions between the end and center.

In considering rational behavior in the last block of five trials, only 8 of the 40 *S*s showed no rational predictions for any of the horses (3 predicted rationally for all five horses, 4 for four, 5 for three, 6 for two, and 14 for one). This group of *S*s was significantly better educated—$M = 14.38$ yr., $t\ (38) = 13$, $p < .001$—than the group that showed rational behavior. This group had slightly less matching behavior, in spite of the fact that with less rational predictions they had more opportunity for matching predictions. Whether this was related to assumed sophistication or some other variable is not clear.

It is of interest to note that the Postdict data do not closely match objective probabilities. The effects of position of horses, Motivation, and Preferred Side singly and/or in combination affect the probability matching. Whether or not these effects are independent of *S*'s active participation in the experiment cannot be determined from this experiment. One thing that is clear is that *S*s' postdictions of events are relatively independent of those factors that lead to differential prediction of Win-vs.-Lose events, as well as the actual predict responses themselves.

One can contrast the analysis of the

generalization responses which was used in all previous studies with the analysis of E_1 responses used for the first time in this study. The E_1 analysis proved to be the more sensitive by having more significant relationships and at the same time including all the significant relationships found in the generalization analysis. In other words, variables and combinations of variables which do not significantly affect the generalization curve may nevertheless significantly affect the E_1 behavior for each horse.

As a final point, one might note that while the probability learning of two consecutive problems seems to occur independently when *S* makes two guesses following a single signal light (Brody, 1958), the results of the present and previous studies suggest that this is not the case when each guess follows its own signal light. The effects of spatial generalization, relatively rapid learning for each horse, the high incidence of rational behavior, and the experimentally produced position preference are all evidence of a lack of independence.

REFERENCES

Brody, A. L. Independence in the learning of two consecutive responses per trial. *J. exp. Psychol.*, 1958, **56**, 16–20.

Brown, J. S., Clark, F. R., & Stein, R. A. A new technique for studying spatial generalization with voluntary responses. *J. exp. Psychol.*, 1958, **55**, 359–362.

Dixon, B. D., & Wickens, D. D. Supplementary Report: Spatial generalization of voluntary responses under two techniques of study and two levels of anxiety. *J. exp. Psychol.*, 1961, **61**, 508–509.

Estes, W. K. Learning theory. *Annu. Rev. Psychol.*, 1962, **13**, 131–132.

Evans, W. O. Two factors affecting stimulus generalization on a spatial dimension. *J. exp. Psychol.*, 1961, **61**, 142–149.

Fager, R. E., & Knopf, I. J. Relationship of manifest anxiety to stimulus generalization. *J. abnorm. soc. Psychol.*, 1958, **57**, 125–126.

Siegel, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two choice uncertain outcome situation. *Psychometrika*, 1959, **24**, 303–316.

Siegel, S. Decision making and learning under varying conditions of reinforcement. *Ann. N. Y. Acad. Sci.*, 1961, **89**, 768–783.

(Received May 23, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 352–359

EFFECT OF UCS INTENSITY ON THE ACQUISITION AND EXTINCTION OF AN AVOIDANCE RESPONSE[1]

K. E. MOYER AND JAMES H. KORN
Carnegie Institute of Technology

123 naive female albino rats were given 30 trials a day for 4 days on a shuttle-box avoidance task. Different groups received 0.5, 1.0, 1.5, 2.5, 3.0, 3.5, and 4.5 ma. of shock during training. The results showed that relatively intense shock interferes with the acquisition of both escape and avoidance responses. Learning falls off rapidly after 1.0 ma. and is probably maximally disrupted at about 2.5 ma. with further increase in intensity having relatively limited additional effect up to 4.5 ma., the maximum intensity used. The effects on extinction were highly variable and the differences between the groups were not statistically significant. However, differences in the manner of extinction as a function of shock level were observed. The discrepancy between these findings and those of previous investigations is discussed.

Studies of the effects of shock intensity on aversive learning which have used rats as *S*s have generally shown that learning increases as a function of shock magnitude (Campbell & Kraeling, 1953; Denenberg, 1959; Dunlap, Gentry, & Zeigler, 1931). Kimble (1955) obtained this result in a wheel-turning avoidance-training apparatus with four shock levels ranging from 0.2 to 2.0 ma. The results of Boren, Sidman, and Herrnstein (1959) were similar. They used eight shock intensities from 0.1 to 3.7 ma. and the rats were required to press a lever to postpone shock. Annau and Kamin (1961) using the conditioned suppression of a bar-pressing response also found that learning increased as the shock intensity increased.

Brush (1957) studied dogs in a shuttle-box avoidance situation using five UCS intensities ranging from 0.34 to 5.78 ma. He used six different indexes of acquisition of the shuttle response and found no significant differences due to shock intensity. He did find that the percentage of *S*s learning the response increased directly as a function of shock magnitude, and the latency of the first escape response decreased as a function of shock magnitude. He found two indications that the relationship between shock intensity and learning might be curvilinear. The latency of the last shock trial showed a significant **U**-shaped trend across the shock-intensity groups as did the mean speed of response during the extinction trials. Trapold and Fowler (1960) found a nonmonotonic relationship between shock intensity and starting speed for the escape response in a runway. However, running speed and shock intensity were *monotonically* related.

During the past several years in this laboratory the behavior of rats subjected to relatively high levels of shock has been observed in a variety of situations and the resulting considerable behavior disorganization has

---

[1] This investigation was supported in part by a research grant M-1298 from the National Institutes of Health, Public Health Service to the senior author.

The authors are grateful to Alan Patz who rendered invaluable assistance in running many of the animals.

been impressive. Also, there is reason to believe from other studies using nonshock motivation (Kaplan, 1952) and from theoretical positions (Easterbrook, 1959; Malmo, 1958) that measures of performance and learning may decrease when the motivational state is raised beyond some optimum level.

The present study explores the relationship between UCS intensity and aversive learning in the shuttle box.

METHOD

Subjects.—One hundred and twenty-three naive albino female rats from the colony of the Carnegie Institute of Technology were used. All *Ss* were between 90 and 110 days of age at the beginning of the experiment.

Apparatus.—The avoidance box consisted of an alley $2\frac{1}{2}$ ft. long × 10 in. wide. The sides of the alley were 16 in. high. It was divided into equal compartments by a $\frac{1}{4}$-in. Masonite partition which was suspended across the alley 5 in. above the floor. The interior of the box was painted battle-ship gray. The floor of each compartment consisted of a grid made of $\frac{1}{8}$-in. brass rods spaced $\frac{3}{8}$ in. apart on center. The grids were hinged at the ends and suspended by springs in the center so that the weight of the rat on the grid activated a microswitch. An Applegate stimulator, Model 228, was used to charge the grids. The top of the alley was covered with $\frac{1}{8}$-in. Plexiglas. A $7\frac{1}{2}$-w. frosted light bulb was hung 24 in. above the center of each grid. A 5-in. speaker was mounted on the apparatus 7 in. behind the center of the box. A 600-cycle tone was generated by a Hewlett-Packard sound generator. The sound level at the grid was approximately 73 db. A control apparatus located in another room automatically timed and presented the events in the avoidance box and recorded the latencies and types of responses made by *Ss*.

Procedure.—Nine groups of *Ss* were run in this experiment. The last three groups were run as a result of the findings on the first six. For convenience in exposition, the first six groups are referred to as Phase 1 and the last three as Phase 2 of the experiment. In Phase 1 there were 15 *Ss* in each group except for Group II which contained 18 *Ss*. In Phase 2 there were 10 *Ss* in each group. The procedure for all *Ss* was exactly the same; it differed only in the magnitude of the shock on the grid during learning. The current magnitudes were: Group I = 0.5 ma., II = 1.0 ma., III = 1.5 ma., IV = 2.5 ma., V = 3.5 ma., VI = 4.5 ma., VII = 2.5 ma., VIII = 3.0 ma., and IX = 3.5 ma.

At the beginning of the experiment all *Ss* were placed in individual cages and given food and water ad lib. Five days later avoidance training was begun. At the beginning of training, the light over the right-hand grid was on. The *S* was placed on the lighted grid facing away from the dark grid. On the first trial, the tone sounded, the light over the right grid went out and at the same time the light on the left grid went on. After a 5-sec. interval, the dark grid was charged with the appropriate current. Shock remained on the dark grid until the beginning of the next trial. The tone was terminated as soon as *S* ran to the lighted grid. On each subsequent trial, the light shifted, the tone sounded, and the entire grid was free of shock for a 5-sec. period followed by shock on the dark grid. The intertrial interval was varied from 10 to 30 sec. unsystematically with a mean interval of 20 sec. Each *S* received 30 trials per day for 4 days.

An avoidance response was defined as a response of crossing from the dark grid to the lighted one within 5 sec. of the presentation of the tone and the changing of the lights. On such a response, *S* did not receive shock. If *S* delayed responding for more than 5 sec., it received shock and the response was defined as escape. Spontaneous crossings (i.e., be-

TABLE 1

MEANS AND *SD*s OF PERCENTAGE AVOIDANCE RESPONSES AND LATENCIES IN PHASE 1 AND PHASE 2

Group	% Avoidances		Latencies	
	M	*SD*	*M*	*SD*
Phase 1				
I—0.5 ma.	74.61	10.34	2.76	.58
II—1.0 ma.	74.58	11.51	2.52	.55
III—1.5 ma.	57.49	18.48	3.44	.86
IV—2.5 ma.	37.85	21.58	4.52	1.13
V—3.5 ma.	58.99	17.69	3.44	.86
VI—4.5 ma.	41.88	17.05	4.23	.86
Phase 2				
VII—2.5 ma.	49.75	18.54	4.25	.90
VIII—3.0 ma.	37.75	21.18	4.36	.94
IX—3.5 ma.	36.25	17.85	3.93	.82

tween trials) dropped out quickly because the dark grid was always charged between trials.

On the fifth day of training the extinction procedure was begun. During the extinction trials no shock was used. The CS was presented as usual but if *S* did not respond within 10 sec. the tone was turned off and the light returned to *S*'s side of the box. When *S* failed to respond three times in a given 10 trials, it met the extinction criterion. It also met the extinction criterion if it made five spontaneous crossings in any 10 trials. The spontaneous-crossing criterion was used because extensive pilot work had shown that during extinction some *S*s lost all fear of the tone and wandered around the box in typical exploratory behavior regardless of whether the tone was on or not. The apparatus was so wired during both acquisition and extinction that the weight of the rat on the dark grid activated the tone. Thus, a spontaneous crossing during extinction meant that the tone

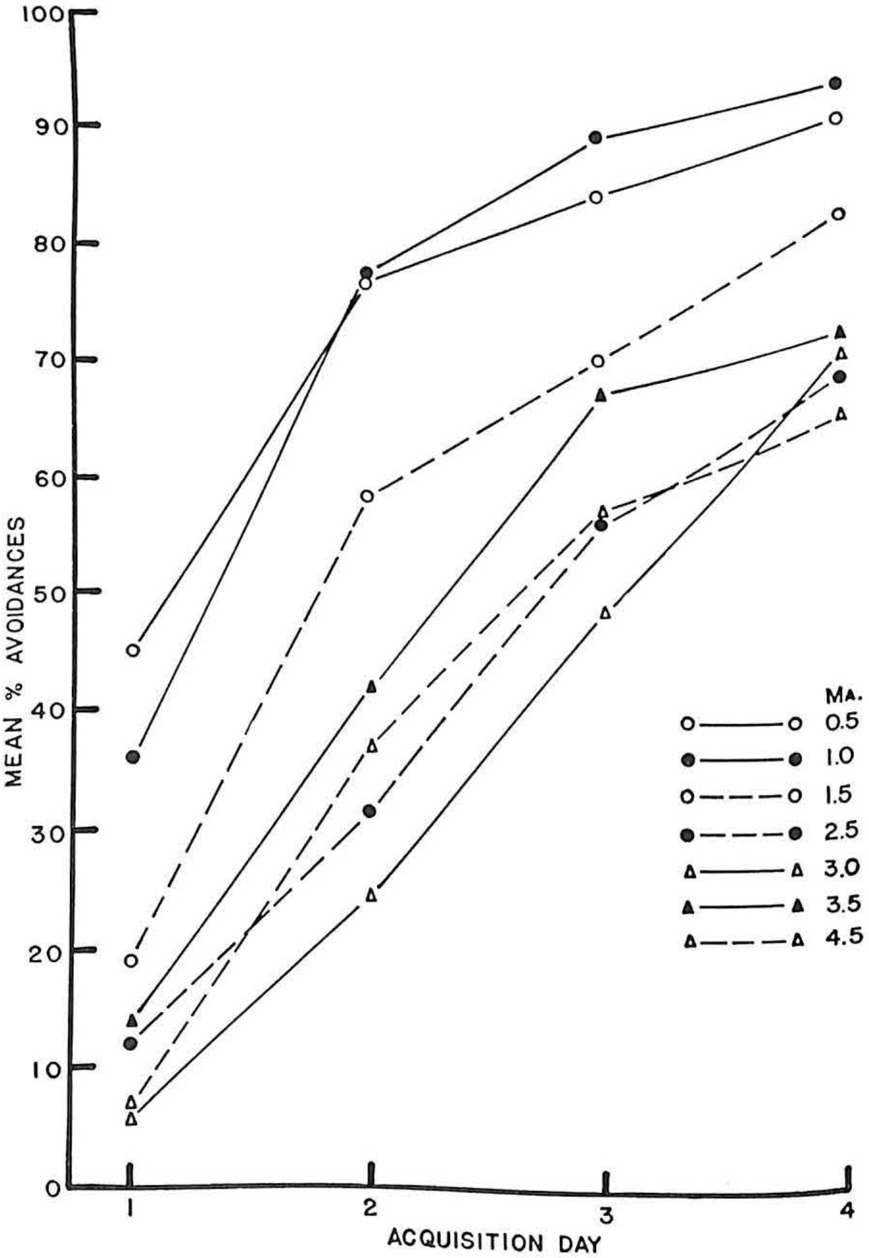


FIG. 1. Mean percentage avoidance responses as a function of acquisition day.

was on all of the time that the rat was on the dark (formerly shock) portion of the grid.

RESULTS

The means and standard deviations of percentage avoidance responses and avoidance latencies for the two phases of the experiment are presented in Table 1. Analyses of variance carried out in the Phase 1 data alone proved significant both for percentage avoidance responses, $F\ (5, 87) = 13.17$, $p < .001$; and for avoidance latencies, $F\ (5, 87) = 13.68$, $p < .001$. These results, along with the t tests performed on all combinations of treatment means, indicated that the curve relating avoidance learning to shock intensity was a complex one which had an inflection point around 2.5 ma., followed by another inflection around 3.5 ma. Although such a complex relationship was possible, no available evidence or theory would predict it. In order to check for the possibility that this result was due to chance, Phase 2 was carried out in an attempt to reproduce this finding. In this phase of the experiment, the 2.5-ma. and 3.5-ma. groups were replicated and the 3.0-ma. group was included to further specify that portion of the curve. Analyses of variance on Phase 2 alone were not significant, either for percentage avoidance responses, $F\ (2, 29) = 1.33, p > .20$; or for avoidance latencies, $F\ (2, 29) = 1.95$, $p > .10$. In all subsequent analyses, the results of Phases 1 and 2 have been combined.

An analysis of variance of the total percentage avoidance responses for all groups with the successive days of acquisition as a second variable indicated significant effects of shock level, $F\ (6, 116) = 10.83$, $p < .001$; days, $F\ (3, 348) = 407.87$, $p < .001$; and interaction of shock level with days, $F\ (18, 348) = 2.35$, $p < .01$. The source of the significant interaction is shown in Fig. 1 where it can be seen that the inflection points in

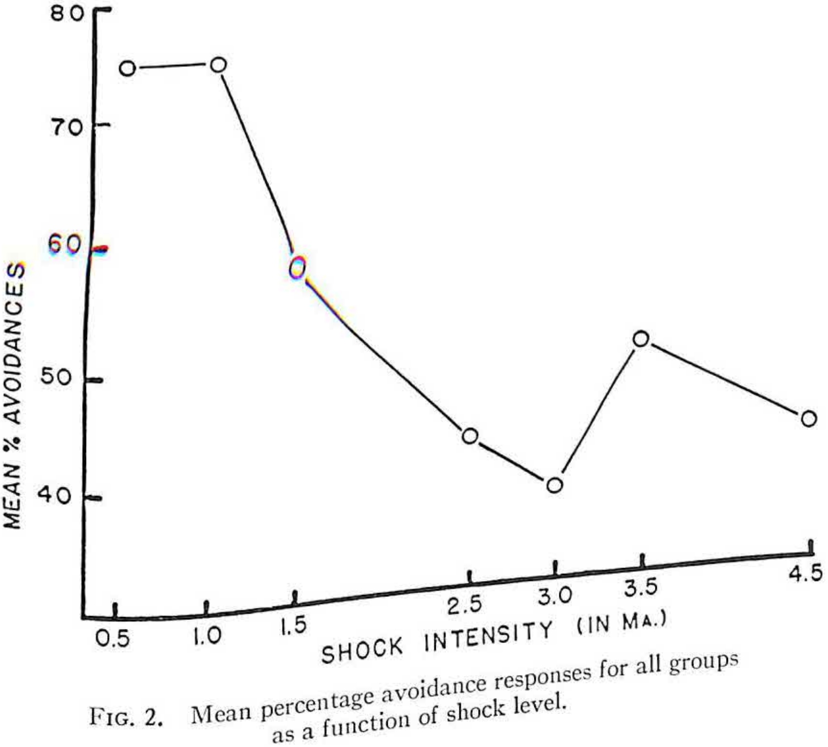


FIG. 2. Mean percentage avoidance responses for all groups as a function of shock level.

TABLE 2

t TESTS OF DIFFERENCES BETWEEN GROUP MEANS FOR PERCENTAGE AVOIDANCE RESPONSES (CELL VALUES ARE ABSOLUTE DIFFERENCES)

	1.0	1.5	2.5	3.0	3.5	4.5
0.5	.03	17.12*	32.01**	36.86**	24.72**	32.73**
1.0		17.09*	31.98**	36.83**	24.69**	32.70**
1.5			14.89*	19.74**	7.60	15.61*
2.5				4.85	7.29	.72
3.0					12.41	4.13
3.5						8.01

* $p < .05$.
** $p < .01$.

the acquisition curves of the low-shock groups occur earlier than the inflections for the high-shock groups.

In Fig. 2, the curve showing percentage avoidance responses as a function of shock intensity is presented. Paired *t* tests were performed on the group means represented by the points on this curve, the results of which are given in Table 2. These tests indicate that the curve decreases from 1.0 ma. to 2.5 ma., with increases in shock intensity beyond this point having no significant effect.

The analysis of variance of response latencies also revealed significant differences between shock levels, $F\ (6, 116) = 11.19, p < .001$; acquisition days, $F(3,348) = 432.64, p < .001$; and the interaction of the variables, $F\ (18, 348) = 1.92, p < .05$.

The curve presented in Fig. 3 relates shock intensity and avoidance latency. The results of the *t* tests

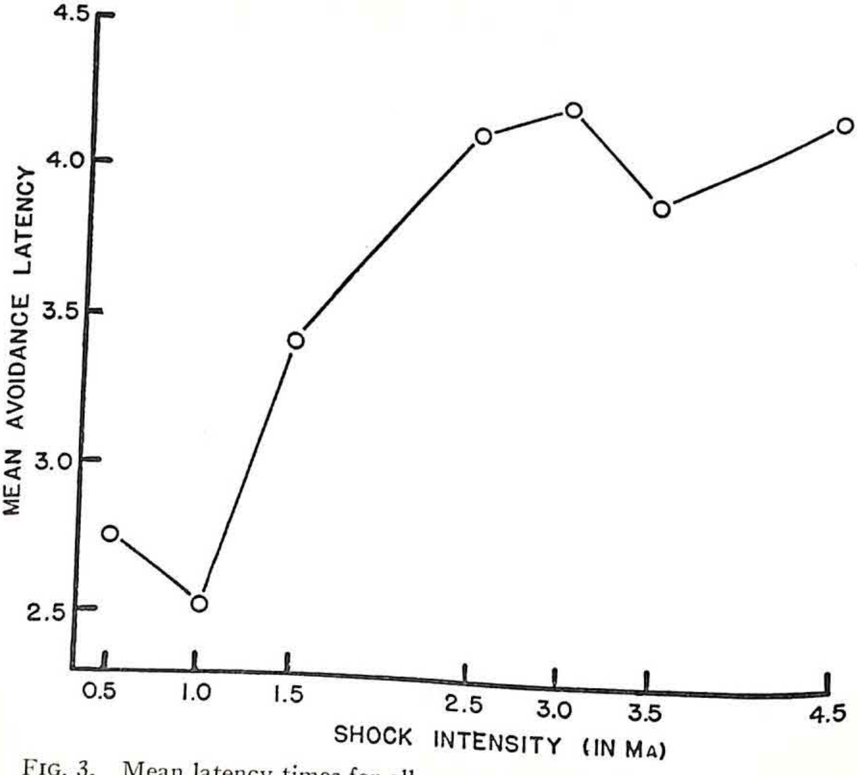


FIG. 3. Mean latency times for all groups as a function of shock level.

TABLE 3

t TESTS OF DIFFERENCES BETWEEN GROUP MEANS FOR LATENCIES (CELL VALUES ARE ABSOLUTE DIFFERENCES)

	1.0	1.5	2.5	3.0	3.5	4.5
0.5	.24	.68*	1.49**	1.60**	1.17**	1.47**
1.0		.92**	1.73**	1.84**	1.41**	1.71**
1.5			.81**	.92*	.49	.79*
2.5				.11	.32	.02
3.0					.43	.13
3.5						.30

* $p < .05$.
** $p < .01$.

TABLE 4

t TESTS OF DIFFERENCES BETWEEN GROUP MEANS FOR LATENCY ON FIRST RESPONSE TO SHOCK (CELL VALUES ARE ABSOLUTE DIFFERENCES)

	1.0	1.5	2.5	3.0	3.5	4.5
0.5	1.13**	.90**	.55**	.58**	.09	.35*
1.0		.23	.58**	.61**	1.22**	.78**
1.5			.35**	.38**	.99**	.55**
2.5				.03	.64**	.20
3.0					.67**	.23
3.5						.44**

* $p < .05$.
** $p < .01$.

carried out on these data are shown in Table 3 and, in general, support the results obtained with total percentage avoidance responses as the measure of avoidance learning.

If 80% avoidance responses on Day 4 is used as an arbitrary criterion of learning the results show that the percentage of *S*s in each group reaching criterion was as follows in ascending order of shock intensity: 93, 94, 73, 52, 60, 56, 40%.

The means for the latency of the first response to shock are plotted in Fig. 4. An analysis of variance showed the treatment means to be significantly different; $F\ (6, 116) = 2.50$, $p < .05$. The results of the paired *t* tests performed on these means are presented in Table 4.

The extinction data, in terms of trials to criterion, were highly variable. An analysis of variance applied to these data showed that the differences between groups were not statistically significant; $F\ (6, 116) = .80$. Because of the great variability in the

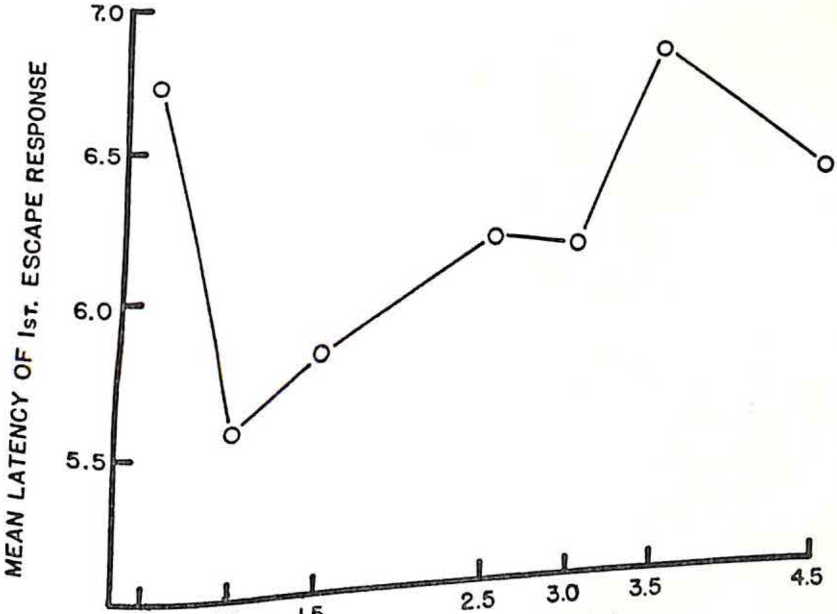


FIG. 4. Mean latency for first response to shock for all groups as a function of shock level.

TABLE 5

Median Trials to Extinction under No-Response (NR) and Spontaneous-Crossing (SC) Criteria

Shock (Ma.)	NR	SC	Combined
0.5	3.5	23.0	15.0
1.0	4.0	23.3	16.0
1.5	4.0	29.5	16.0
2.5	3.0	55.5	3.7
3.0	3.0	14.5	3.5
3.5	4.0	47.5	4.4
4.5	3.0	79.0	3.7

criterion measures, the nonparametric Kruskal-Wallis analysis of variance was also performed on the data and this too failed to attain an acceptable level of significance; H (6) = 7.37, $p < .20$.

It will be recalled that two extinction criteria were used in this experiment: a "no-response" criterion and a "spontaneous-crossing" criterion. The median number of trials to each of these criteria and to the two combined are given in Table 5. It can be seen that in all groups, many more trials are required to extinguish by the spontaneous-crossing criterion than by the no-response criterion. Also, more trials to extinction are required by the low-shock groups than by the high-shock groups.

This last result is probably due to the fact that as shock intensity increases, the percentage of *S*s in each group that extinguish by the no-response criterion also increases: 40, 39, 64, 80, 84, 87%. A chi-square test applied to the frequency data (not the percentages) proved to be significant, χ^2 (6) = 19.65, $p < .01$, indicating that there are differences in the manner of extinction as a function of shock level.

Discussion

When Phases 1 and 2 are combined, the general trend of the results is quite clear. Relatively intense shock interferes with the acquisition of both escape and avoidance responses. Learning falls off rapidly after 1.0 ma. and is probably maximally disrupted at about 2.5 ma. with further increases in intensity having relatively limited additional effect up to our maximum intensity of 4.5 ma. Tables 2 and 3 show that the last four points on the curve do not differ significantly. The unusually high performance of the 3.5-ma. group in Phase 1 was not reproduced in Phase 2, thus confirming our suspicion that it was a chance result.

Differences in rate of acquisition were indicated by the significant Treatments × Days interactions in the analyses of variance done on percentage avoidance responses and response latencies. The groups which performed best over all 4 days (0.5 ma., 1.0 ma., 1.5 ma.) also showed inflection points in their learning curves sooner than did the groups with poorer overall performance. That is, the low-shock groups perform better than the high-shock groups throughout the experiment and also show faster rates of acquisition of the avoidance response.

These findings differ considerably from those of Annau and Kamin (1961), Boren, Sidman, and Herrnstein (1959), and Kimble (1955). However, they studied quite different responses. In the CER used by Annau and Kamin (1961) the measure of learning is the disruption of ongoing behavior. There is no instrumental escape or avoidance response to be disrupted by the intense shock. In the Boren, Sidman, and Herrnstein (1959) study the response of bar pressing to avoid or escape shock was highly overlearned before *S*s were exposed to the higher shock intensities. They had been pretrained for 100 hr. at what we found to be slightly greater than optimal intensity, 1.2 ma. Their study actually explored the effect of various shock intensities on a highly overlearned response. Kimble used a wheel-turning avoidance response, and it may simply be the case that wheel turning is less susceptible to disruption by shock than the shuttle type of response. Since 2.0 ma. was the highest shock intensity used by Kimble, it is impossible to determine whether or not the latency curve he presents would have taken an upswing at higher intensities.

Brush (1957) using a shuttle-box avoidance with dogs, obtained evidence for a disruption of behavior with intense shock. Although the procedures differ considerably, it is interesting to note that rats appear to be much more susceptible to shock than are dogs. Brush got optimum learning (as measured by percentage of *Ss* meeting criterion) at 4.8 ma.; whereas in this study optimum learning with rats occurred at about 1.0 ma. using a similar criterion. It may be that Brush would have found a greater disruption of learning had he used higher intensities.

In order to consider the effect of shock level on performance prior to learning, the latency of the first response to shock was examined. Figure 4 clearly shows that the performance is relatively poor at the lowest shock level, reaches optimum at about 1.0 ma., and then gradually decreases again.

Observations indicate that the .5-ma. shock level disturbs *Ss* but does not activate them to intense running activity as do the higher shock levels. Those shock levels in the middle range tend to elicit relatively pure running behavior, while the highest levels result in considerable disorganized behavior including bar biting, jumping, falling, and bumping into walls. Thus, the middle shock range generally produces behavior which leads most quickly to escape from shock in the shuttle-box situation. In other situations where different types of escape responses are required, different shock levels might produce optimum performance.

Like Kimble's (1955) study, this experiment showed that shock intensity during learning had no significant effect on number of trials to extinction. However, significantly more *Ss* in the high-shock groups extinguished by the "no-response" method than by the "spontaneous-crossing" method. The typical extinction behavior of *S* extinguishing by the no-response method was freezing at the tone onset, whereas, those *Ss* extinguishing by the spontaneous-crossing method were active in the box and in response to the tone. Freezing is a very common innate fear response in the rat. Thus, it is possible that although the high-shock *Ss* did not learn the running response to the tone as well as *Ss* in the lower shock groups, they may have learned a fear response to the tone better. Further research is needed to clarify this point.

REFERENCES

ANNAU, Z., & KAMIN, L. J. The conditioned emotional response as a function of intensity of the US. *J. comp. physiol. Psychol.*, 1961, **54**, 428–432.

BOREN, J. J., SIDMAN, M., & HERRNSTEIN, R. J. Avoidance, escape, and extinction as functions of shock intensity. *J. comp. physiol. Psychol.*, 1959, **52**, 420–425.

BRUSH, F. R. The effects of shock intensity on the acquisition and extinction of an avoidance response in dogs. *J. comp. physiol. Psychol.*, 1957, **50**, 547–552.

CAMPBELL, B. A., & KRAELING, D. Response strength as a function of drive level and amount of drive reduction. *J. exp. Psychol.*, 1953, **45**, 95–101.

DENENBERG, V. H. Interactive effects of infantile and adult shock levels upon learning. *Psychol. Rep.*, 1959, **5**, 357–364.

DUNLAP, K., GENTRY, E., & ZEIGLER, T. W. The behavior of white rats under food and electric shock stimulation. *J. comp. Psychol.*, 1931, **12**, 371–378.

EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.*, 1959, **66**, 183–201.

KAPLAN, M. The effects of noxious stimulus intensity and duration during intermittent reinforcement of escape behavior. *J. comp. physiol. Psychol.*, 1952, **45**, 538–549.

KIMBLE, G. A. Shock intensity and avoidance learning. *J. comp. physiol. Psychol.*, 1955, **48**, 281–284.

MALMO, R. B. Measurement of drive: An unsolved problem in psychology. In Marshall B. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958.

TRAPOLD, M. A., & FOWLER, H. Instrumental escape performance as a function of the intensity of noxious stimulation. *J. exp. Psychol.*, 1960, **60**, 323–326.

(Received May 27, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 360–364

CONCEPT ACQUISITION AS A FUNCTION OF STIMULUS-EQUIVALENCE PRETRAINING WITH IDENTICAL AND DISSIMILAR STIMULI[1]

RICHARD D. PETRE

George Peabody College for Teachers

A matching-to-sample concept-learning task was given to 4 experimental groups and 1 control group of 25 undergraduates each. The matching was on the basis of size (large or small) and linearity (circles and crescents or squares and triangles). The experimental groups were given a stimulus-equivalence pretraining task. 2 of the groups were trained with stimuli identical to those used in the concept task by operations intended to strengthen mediating responses that would facilitate task performance for one group, and interfere with performance for the other. The other 2 experimental groups were given similar training tasks except that the stimuli used were not identical with the concept-task stimuli. The prediction that the facilitating and interfering effects would be less pronounced for the latter 2 groups was generally confirmed.

Several recent investigators have suggested that a single-unit S-R theory does not adequately represent adult human concept learning (Kendler, Glucksberg, & Keston, 1961; Kendler & Kendler, 1962; Kendler & Vineberg, 1954; Mednick & Freedman, 1960). An alternative model focuses on implicit stimulus-response events which are assumed to mediate between the external stimulus and the overt response (Kendler & Kendler, 1962). The inferred process is conceptualized as a complex behavioral chain that differs from conditioning only in its degree of complexity (Kendler et al., 1961).

A basic assumption of the mediational model is that the inferred, mediating stimulus-response relationships obey the same principles that operate in observable S-R relationships (Kendler & Kendler, 1962). Mednick and Freedman (1960) hypothesized that ease of attainment of a concept is directly related to its dominance level. Increasing the strength of association between concept instances and the appropriate implicit concept response should facilitate its acquisition.

The aim of the present study was to influence performance on a concept-learning task by operations designed to establish mediating responses that facilitate or interfere with acquisition. The problem used was a four-choice matching to a sample compound-concept-learning task. The correct choice was the same size (large or small) and the same shape (rectilinear or curvilinear) as the sample stimulus. It was expected that if the rectilinear stimuli (triangle and square) and the curvilinear stimuli (crescent and circle) were each conditioned to a common nonsense syllable, subsequent performance on the concept-learning task would be facilitated due to the strengthening of mediating responses appropriate to the concept task. On the other hand, if one of the rectilinear figures and one of the curvilinear figures were conditioned to a common nonsense syllable, and the other member of each pair to a different nonsense syllable, performance would be lower because the mediating responses presumed to have been strengthened would not be appropriate to the problem.

---

[1] This research was supported in part by Grant 2M-6480 from the National Institute of Mental Health.

A second concern of the study was with the relationship between the stimuli used in stimulus-equivalence pretraining and the stimuli used in the concept-learning task. In the pretraining phase the stimuli used in conditioning could be identical to those used in the concept problem or they could be physically different, but similar in the respect that they could be classified as to curvilinear-rectilinear. In the first case, that of training with identical stimuli, the facilitation or interference could be interpreted as the strengthening of a response habit that is directly, positively or negatively, transferable to a new learning situation. In the second case, that of training with physically dissimilar stimuli, the effects should be less pronounced than in the first case due to generalization decrement. Three predictions were made. First, it was expected that stimulus-equivalence pretraining would facilitate performance on the concept-learning task when stimuli of a common linearity value were conditioned to a common nonsense syllable. Second, when stimulus pairs differing in linearity are conditioned to the same nonsense syllable, subsequent performance should be lowered. Third, the effects of pretraining with stimuli identical to those used in the learning task should be more pronounced than in the case of pretraining with physically dissimilar stimuli.

METHOD

Subjects.—The *S*s were 125 volunteers enrolled in undergraduate psychology courses in which credit was given for participating in psychological experiments.

Apparatus.—The concept-learning task was located in a $7 \times 7 \times 6\frac{1}{2}$ ft. room, illuminated by a shaded, 60-w. bulb. An electric exhaust fan located above the ceiling provided background noise sufficient to mask equipment sounds. A set of five miniature projectors $1\frac{1}{2} \times 2$ in., was arranged on an oblique, 16×19 in. metal display panel. The sample stimulus presentation projector was centered, about 6 in. from the top. A small green light, which served as a reinforcement signal was mounted directly below the top projector. Four projectors, identical to the sample stimulus projector, were arranged in a row below the reinforcer light. A push-button switch was mounted below each of the four bottom projectors. The panel was placed on a 30-in. high table at which *S* was seated. The sequence of stimuli presented was programed, and responses were recorded on a running tape recorder. The electronic apparatus and the recorder were located in an adjoining room.

Procedure.—The *S*s were assigned to one of five groups on the basis of the order in which they arrived to participate in the experiment. There were four experimental groups and one control group with 25 *S*s each. The procedure consisted of two phases. In the first phase the experimental groups were given different tasks which consisted of learning to respond to geometrical figures by saying nonsense syllables. The training task was performed in the room in which the concept-learning task subsequently took place. For two of the experimental groups a set of four, 3-in. square cards containing a square, an equilateral triangle, a circle, and a crescent, was used. The line drawings were 1 in. in height. These two groups will be referred to as the direct facilitation (DF) and the direct interference (DI) groups. The DF *S*s were seated in the experimental room and given a card with BIC and SAR printed on it. They were instructed to say either BIC or SAR to each geometrical figure they would be presented, and that *E* would tell them if they were right or wrong each time. The *E* responded "right" when BIC was said to the square and the triangle and SAR to the circle and crescent. Any other combination was called "wrong." The cards were presented in random order until eight consecutive correct responses were made by *S*. The DI *S*s were given the same instructions, but were reinforced when BIC was said to the square and the circle, and SAR to the triangle and crescent. The other two experimental groups will be referred to as the generalized facilitation (GF) and the generalized interference (GI) groups. A different set of cards was used in the pretraining phase with these groups. These were two curvilinear figures: a reversed **S** and a double-convex lens; and two figures composed of straight lines: a vertical line bisected by a diagonal line of equal length so as to make a smaller angle of 45°, and an inverted **T**. The line drawings were 1 in. high, and centered on 3-in. square cards. The instructions to *S*s in these groups were the same as those for the DF and DI *S*s. The GF *S*s were reinforced for saying BIC to the two rectilinear figures and SAR to the two curvilinear figures; while the GI *S*s learned to say BIC to the bisected line and the double convex lens figures and SAR to the inverted **T** and reversed **S** figures. All *S*s learned to the same criterion of eight consecutive correct responses. The control group (C) received no treatment in the stimulus-equivalence training phase.

The second, or testing phase, consisted of learning a compound concept problem. All *S*s performed the same task. The task involved learning to match one of four stimuli to a sample stimulus on the basis of size and linearity. Thus, if the sample stimulus was a small triangle, the correct choice would be a small rectilinear figure. The stimuli varied in terms of four forms (circle, crescent, square, and triangle) and two sizes. The stimuli appeared on the projectors as solid white forms on a black background and were either 1 in. or 1½ in. high. The *S*s were instructed to push the button corresponding to the stimulus with which they thought the sample stimulus should go, and if the choice was correct the green light would be turned on. The stimuli remained until a response was made. If the response was correct the green light came on for 3 sec. and 2 sec. later the next set of stimuli was presented. If the response was incorrect there was a 5-sec. interstimulus interval. The stimuli were presented in 12 different combinations, and the 12 combinations were presented in four different permutations for a total of 48 trials.

If the learning criterion of 10 successive correct trials was not reached, the same program was presented again. If the learning criterion was not reached in 96 trials, the *S* was eliminated. Three *S*s were eliminated because they did not reach the criterion in the allowed number of trials.

TABLE 1

MEANS AND *SD*s OF TRIALS TO CRITERION FOR PHASE I LEARNING

Group	*M*	*SD*
DF	6.84	4.93
GF	7.48	4.48
GI	7.28	5.01
DI	8.04	5.73

In each set of response-choice stimuli there was one of each of the four forms. On each trial the correct choice was a different figure from the sample stimulus. A sample trial would be a small triangle appearing on the stimulus sample projector with a small square, a large crescent, a small circle, and a large triangle appearing on the response-choice projectors. The correct response would be pushing the button under the small square. The reinforced position was varied randomly over trials.

RESULTS

Table 1 presents means and *SD*s for trials to criterion for the first phase of the experiment. There were no differences between the groups in speed of learning, $F\ (3, 96) = .24$, $p > .50$. Acquisition curves for the

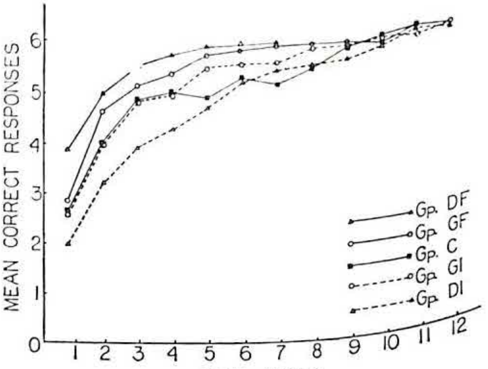


FIG. 1. Mean correct responses during acquisition of the Phase II concept problem in blocks of six trials each.

TABLE 2

Means and *SD*s of Errors to Criterion for the Concept-Learning Task

Group	*M*	*SD*
DF	3.88	4.08
GF	6.56	7.35
C	10.64	12.91
GI	9.32	9.82
DI	14.24	12.04

five groups are presented in Fig. 1. Each *S* was assumed to have made 100% correct responses after reaching the learning criterion, so each point represents the entire group. Group performances were in the predicted direction during the early stage of acquisition, however the fluctuation of means during the later stage results from only two or three *S*s in each group who had not yet reached criterion. The criterion measure used in the statistical analysis was errors to criterion. Group means and *SD*s are presented in Table 2.

The Kruskal-Wallis analysis of variance was used to compare groups, $H\ (4) = 17.25,\ p < .01$. The Mann-Whitney *U* test was used in comparing pairs of groups, and the results are reported in Table 3. With a two-tailed test four of the comparisons reached a significance level of .05.

TABLE 3

Probabilities Associated with the Differences between Pairs of Groups for Errors to Criterion (Two-Tailed Test)

	GF	C	GI	DI
DF	.170	.010	.004	.001
GF		.190	.162	.008
C			.992	.098
GI				.124

Discussion

According to mediational theory the concept problem in this experiment seemed to require at least three processes for solution. First, mediating responses associated with linearity would need to be strengthened, one with rectilinear figures and one with curvilinear figures. Second, a mediating response associated with small figures, and another, with large figures would need to be strengthened. Third, the appropriate combination of mediating responses needs to be elicited. The lower performance of the DI group is explained in terms of the strengthening of inappropriate implicit responses which must be extinguished in the concept-acquisition phase. The observed difference between the GF and GI groups is explained in terms of a stimulus-generalization gradient. The mediating responses which are assumed to have been strengthened in the training phase transfer to other stimuli which have some elements in common with the original conditioned stimulus. Because the stimuli used in the concept-learning phase were not identical to those used in the training phase, the conditioned mediating responses should not be elicited in full strength by the new stimuli. Therefore the performance of the GF *S*s on the concept task should be facilitated, but less than that for the DF group. The same interpretation is made for the GI *S*s.

Although statistical significance was not reached in all comparisons, an S-R mediation theory would seem to have utility in predicting and explaining the observed performances. The role of stimulus generalization seems to require more explication as to its function in transfer-of-training phenomena between concept-learning tasks. For example, if the number of conditioning trials in Phase I were varied it would be predicted that the facilitating or interfering effects would increase for the DF and DI groups with increases in the number of conditioning trials. If the effects increased for the GF and GI groups with increased training trials an interpretation based upon a flattening of the generalization gradient as a function of increased trials would be possible. On the other hand, if transfer effects decreased with increasing trials, an explanation based on a steepening of the generalization gradient might be proposed.

REFERENCES

KENDLER, H. H., GLUCKSBERG, S., & KESTON, R. Perception and mediation in concept learning. *J. exp. Psychol.*, 1961, **61**, 186–191.

KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, **69**, 1–16.

KENDLER, H. H., & VINEBERG, R. The acquisition of compound concepts as a function of previous training. *J. exp. Psychol.*, 1954, **48**, 252–258.

MEDNICK, S. A., & FREEDMAN, J. L. Facilitation of concept formation through mediated generalization. *J. exp. Psychol.*, 1960, **60**, 278–283.

(Received May 31, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 365–372

REACTION TIME AND ALPHA BLOCKING IN NORMAL AND SEVERELY SUBNORMAL SUBJECTS

BEATE M. HERMELIN AND P. H. VENABLES

Medical Research Council Unit, Institute of Psychiatry, London

It has been suggested that stimuli falling into a period of blocked alpha rhythm result in faster reaction times than those which fall into a period when the alpha is present. An experiment is described in which the interval between a warning signal and the reaction-time stimulus was varied irregularly. Thus some RT responses were given when the alpha was still blocked by the forewarning, while for others the alpha had returned. The results did not confirm findings that RT is faster when the stimulus falls into a period in which the alpha is blocked, than when it is not.

Two sets of experimental findings concerning the effect of forewarning on speed of reaction time are relevant to the present experiment. One is that if the length of foreperiod is varied irregularly RT speed is not increased or decreased (e.g., Klemmer, 1956; Woodrow, 1914). When warning signals are presented in blocks using a constant foreperiod, it is found that an interval of about 2 sec. produces the fastest RT. Longer or shorter foreperiods result in increases in RT. The other view is that increased RT speed should result if the forewarning signal blocks the alpha rhythm, and thereby activates the EEG. Lansing, Schwartz, and Lindsley (1959) and Fedio, Mirsky, Smith, and Parry (1961) provided evidence to show that when a stimulus was preceded by a warning signal which resulted in EEG alpha blocking, normal *S*s reacted more quickly than when there was no preparatory signal or when the signal did not produce blocking.

Hermelin and Berkson (1964) have shown a relationship of RT to period of forewarning in mentally deficient *S*s, even when the foreperiod was varied irregularly. Berkson, Hermelin, and O'Connor (1961) have shown that the length of alpha blockade is shorter in defective, than it is in normal *S*s. The use of these patients in the present experiment was therefore expected to throw additional light on the relationship between EEG activation and RT.

In the present experiment foreperiods were varied irregularly. With normal *S*s this should result in RTs which do not differ significantly with foreperiod length. On the other hand, it should be expected, on the basis of the activation hypothesis, that stimuli falling in a period when alpha is blocked (i.e., after a short forewarning) would result in faster RTs than those which fall in a period where alpha has returned (i.e., after a long forewarning). Variation in RT according to length of forewarning would thus be expected. As there is no reason to think that irregularity of presentation of forewarning periods will affect the pattern of alpha blocking the expected findings of the effect of alpha blocking and those of irregular forewarning period are thus opposed.

METHOD

Two conditions were used for forewarning presentation. In the first the forewarning light was switched on for a short period, there then followed a period of darkness before the stimulus sound was presented. In the second condition the forewarning light remained on until 1.5 sec. after the presentation of the sound stimulus. Condition 2 was used to lengthen the period of alpha blockade. This should have resulted in a shortening of the RT to longer forewarning intervals if the findings of Lansing et al. (1959) and Fedio et al. (1961) were correct. A left-occipital bipolar EEG electrode placement was used with an earth electrode over the mastoid process. All *S*s were, as far as could be established, right handed. The *S* was seated and had at his right hand a key which was pressed as a reaction to the stimulus. Auditory RT stimuli and a visual forewarning were used. The loudspeaker for the delivery of the 70-db. white-noise stimulus and the warning light which was a 1,000-ft-L approximate point source were 5 ft. in front of *S* and 1 ft. above his line of sight. The intensity of the light was sufficient to evoke alpha blockade even if the eyes were closed.

The duration of the stimulus noise was .4 sec. In Cond. 1 the forewarning light stimulus lasted .4 sec., in Cond. 2 it lasted 1.5 sec. beyond onset of the noise stimulus. Six forewarning periods were used, they were .5, 1, 2, 4, 6, and 8 sec., timed from the onset of the forewarning light to the onset of the stimulus sound. Reaction times were measured to the nearest millisecond on a dekatron timer. The *S* was in complete darkness, the timing and EEG apparatus were situated in an adjacent room.

Forewarning periods were varied in the order of successive rows of 6 × 6 Latin squares. In each condition 36 reaction times were made followed by a rest of 1 min. and a further 36 reactions. The forewarning stimulus was presented at the time of a train of alpha waves of at least ½ sec. in duration. The times between successive presentations were thus irregular, and were all longer than 7 sec. Conditions 1 and 2 were presented alternately first and second to successive *S*s, and were balanced within groups. Sufficient practice was given to each *S* to enable him to perform the task with few errors. This practice period varied widely between *S*s.

Six normals, six non-Mongol imbecile, and six Mongol imbecile patients took part in the experiment. The mean ages of the three groups were: normals 28.8, *SD* 6.8; imbeciles 25.3, *SD* 2.8; and Mongols 24.0, *SD* 3.8 yr. The mean alpha frequencies for each group were: normals 9.9, *SD* 1.2 cps; imbeciles 9.2, *SD* 1.1 cps; Mongols 9.8, *SD* 1.2 cps. The mean percent time alpha indexes, while resting in a darkened room, were: normals 68.9, *SD* 16.8; imbeciles 50.0, *SD* 23.6; and Mongols 78.5, *SD* 17.4. Intelligence quotients were available for the patient groups; the non-Mongol imbeciles had a mean IQ of 43.6, *SD* 1.8 and the Mongols 38.6, *SD* 3.8. All *S*s were male and were tested twice, as far as possible at the same hour on successive days.

RESULTS

First the EEG records were examined to determine the presence or absence of alpha rhythm at the time of the presentation of each of the 144 stimuli for each *S*. Alpha was considered to be present at the time of the warning signal when there were 8–13 cps waves which preceded the onset of the warning for at least half a second. Alpha was considered to be present after a block at the start of a train of at least ½ sec. of waves of the correct frequency. The amplitude of the largest of these was not to be smaller than the smallest in the ½ sec. preceding the block. When alpha was present at the time of the warning signal, was blocked by it, and remained blocked until the occurrence of the stimulus, the blocking was considered "inclusive" (I); when the alpha had returned before the onset of the stimulus the blocking was "exclusive" (E); when after blocking alpha had returned for at least ½ sec. and then blocked again until the stimulus occurred, this was called a "spontaneous" (S) block. These categories are illustrated in Fig. 1.

In designing the experiment the assumption was made that in Cond. 1 the .4-sec. flash of light followed by a varying period of darkness would result in a limited period of alpha blocking. On the other hand it was expected that in Cond. 2 with the light remaining lit until after the presentation of the noise stimulus the alpha block would be extended. The percentage of I category patterns out of the total of I, E, and S was calculated for each *S* in each time interval and condition. These results are shown graphically in Fig. 2.

Here it can be seen that the percentage of I category patterns in Cond. 1 for all groups fell as the forewarning interval increased. In the data from both defective groups at about 1.7 sec. there was an equal chance of the noise stimulus falling in a period of blocked or unblocked alpha. This figure may be taken as a crude average of the block length for all defective *S*s to a .4-sec. flash of light. On the same basis the comparable figure for normals is about 2.1 sec. In Cond. 2, even at the longest forewarning interval the percentage of I responses for the normal and Mongol groups never fell to less than 50% of the total of I, E, and S categories. On the other hand the percentage of I responses at different foreperiods in the imbecile group in this condition followed very closely those shown in Cond. 1.

An analysis of variance carried out on these data showed the interactions Conditions × Intervals, F (5, 165) $= 9.54$, $p < .001$, and Conditions × Groups, F (2, 165) $= 25.5, p < .001$, to be the only significant terms. As both these terms are well beyond the .001 level the results of this analysis carried out on percentage data may be considered acceptable. These results confirm the differences between the electrophysiological behavior of imbeciles as compared with Mongols and normals as seen in Fig. 2. In all groups in Cond. 1 the number of inclusive blocks decreased sharply with lengthening foreperiod. On the other hand, while imbeciles behaved in Cond. 2 much as they did in Cond. 1, the normal and Mongol *S*s showed a different pattern of alpha blocking. In these *S*s, even with the 8-sec. foreperiod more than half of the RT stimuli fell within a period in which the alpha remained blocked. No explanation can at present be offered for this observed difference in the maintenance of alpha blockade between Mongol and non-Mongol imbeciles.

If the findings of Woodrow (1914) and Klemmer (1956) held, no effect of forewarning interval would be expected in the normal *S*s, although this would be expected to be a significant variable in the defective groups (Hermelin & Berkson, 1964). On the other hand it would be expected that after a short forewarning period the alpha rhythm would be blocked at the onset of the stimulus while it would have returned after a long forewarning period. Thus if the EEG activation hypothesis were correct, a significant

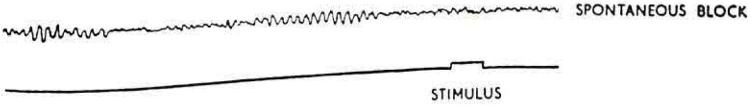


FIG. 1. Examples of records showing inclusive, exclusive, and spontaneous blocks. (Artificially superimposed records from the same *S*.)

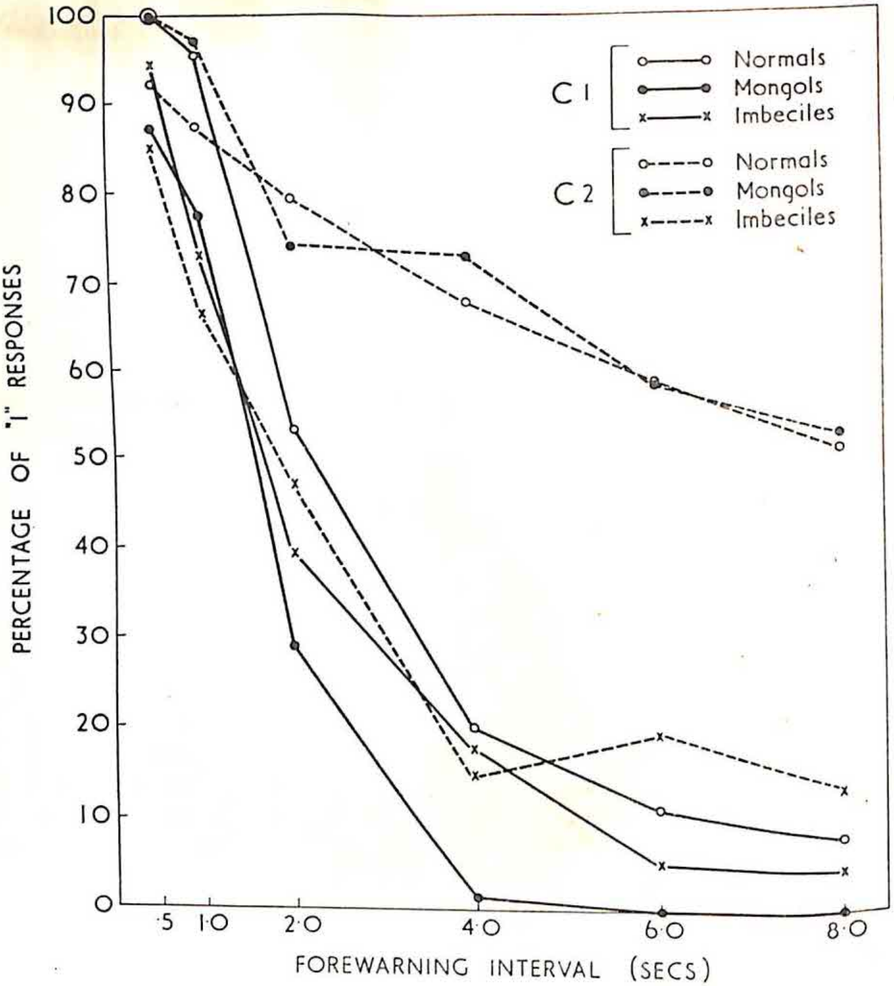


FIG. 2. Percentage of I responses shown by each diagnostic group in two experimental conditions, over six forewarning intervals.

lengthening of RT with forewarning interval would be expected in Cond. 1 in all groups. If the alpha blockade was maintained by the presence of the signal light which was on throughout the forewarning period in Cond. 2 no change in RT with change in foreperiod length would be expected. A significant interaction term in the analysis between conditions and forewarning intervals would then follow.

Analyses of variance which compared the effects of the two conditions and the six forewarning periods were carried out individually for each *S*. Subsequently, three separate group analyses were made. These analyses showed that there was no significant interaction between conditions and intervals for any of the groups—Normals: *F* (5, 643) = 0.99; Imbeciles: *F* (5, 643) = 0.43; Mongols: *F* (5, 421) = 0.51. Thus the results predicted from the activation hypothesis as outlined above were not confirmed. In the absence of a significant Intervals × Conditions interaction term the RTs at the various forewarning intervals for each of the groups are presented in Fig. 3 combined over conditions.

For the normals there was little change about a mean RT of .23 sec. over all intervals. This was confirmed by the absence of a significant intervals term in the analysis of variance. In the imbeciles a highly significant overall term for intervals, *F* (5, 643) = 16.1, $p < .001$, reflected the general trend of the data for individual *S*s. In the Mongols the interval term was also statistically significant, *F* (5, 421) = 3.1, $p < .01$. However, though there was a general increase in RT from 1 to 8 sec. forewarning period, the most marked effect was in the slowness of the RT with the .5-sec. forewarning interval.

Analyses of variance were carried out separately for each *S* on the RT data falling into EEG Categories I, E, and S. This was done only for Cond. 1 as too few RTs were found in the E and S categories in Cond. 2 in the normal and Mongol data. These data are presented in Table 1. There was no significant difference for any individual in the normal and Mongol groups between RTs in different EEG categories. In the imbecile group only one individual's analysis of variance showed that the RT in the

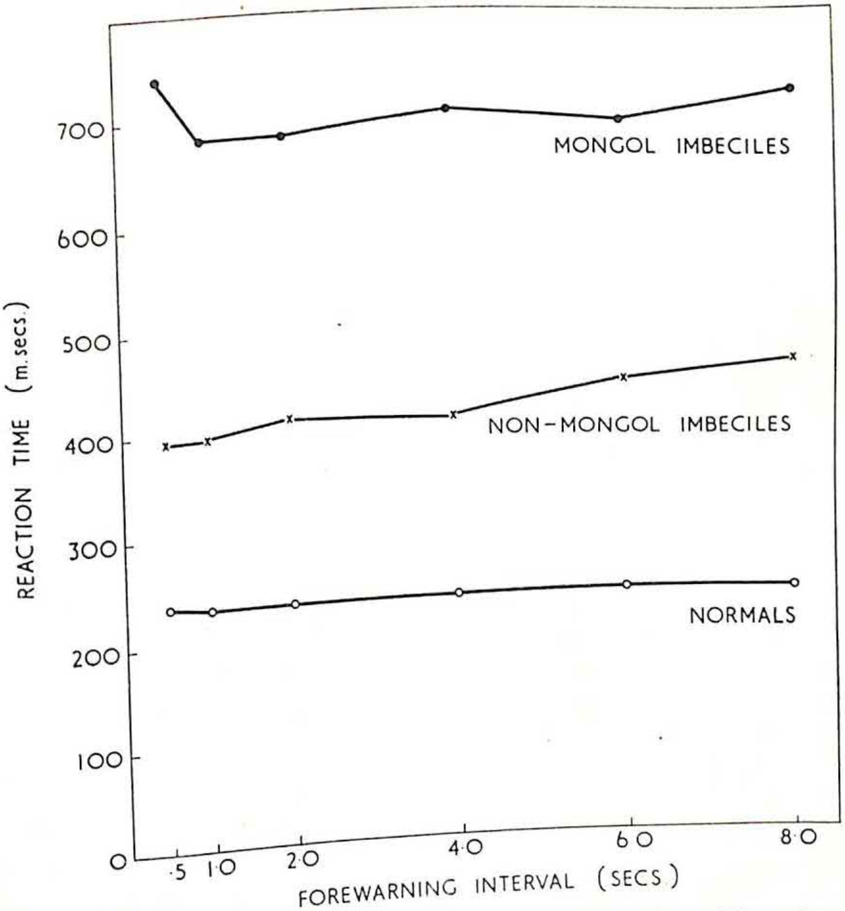


FIG. 3. Mean reaction times of normal, Mongol, and non-Mongol imbecile groups with six forewarning intervals.

TABLE 1

MEAN REACTION TIMES IN MILLISECONDS GIVEN BY NORMAL, IMBECILE, AND MONGOL GROUPS IN THREE CATEGORIES OF EEG ACTIVATION

S No.	Categories of EEG Activation				Significance of Difference between EEG Categories	
	I	E	S	*SD*	*F*	*df*
Normals						
1	185	185	179	29	0.84	2/63
2	258	235	230	41	3.06	2/68
3	198	206	204	26	0.79	2/65
4	189	198	183	30	1.10	2/65
5	307	290	285	75	0.76	2/65
6	276	298	293	48	1.73	2/63
M	236	235	229			
Imbeciles						
1	330	378	371	92	2.11	2/64
2	311	379	362	98	2.99	2/65
3	490	539	610	121	1.23	2/60
4	292	329	347	61	1.79	2/52
5	395	415	510	96	5.46**	2/69
6	333	360	359	81	1.31	2/67
M	358	400	427			
Mongols						
1	813	899	783	260	1.43	2/67
2	920	926	843	149	2.03	2/58
3	531	573	533	132	0.81	2/65
4	794	699	684	203	2.62	2/66
5	346	341	327	96	1.01	2/69
6[a]	—	—	—	—	—	—
M	681	688	630			

[a] The RT data for this *S* were unreliable and were therefore omitted from all analyses.
** $p < .01$.

spontaneous block category was significantly longer than in either of the other categories which did not differ between each other. However, as seen in Table 1, all imbeciles showed a tendency to give shorter RTs when the alpha was blocked than when it was not, though as in the normal and Mongol groups, this difference did not reach the level of statistical significance. In view of this appearance of shorter RT when the alpha rhythm was blocked a further analysis was undertaken.

For any one time interval some RTs were recorded when the alpha was blocked (I) and others when it was not (E). From the data of the six imbecile *S*s, 83 pairs of RTs could be found which were matched for time intervals, conditions, and *S*s. A correlated *t* test between the means of the RTs in the two categories was carried out. A *t* of .72 showed that there was no significant difference between the mean RTs of 359.6, *SD* = 98.4 msec. for I, the alpha-blocked condition, and 369.4, *SD* = 109.3 msec. for E, the alpha-present condition. It is thus evident that the apparent suggestion of differences between I and E categories in the imbeciles was due to those being confounded with time intervals. When matched for this variable the difference was not present.

DISCUSSION

The results presented do not confirm the previous findings of Lansing et al. (1959) and Fedio et al. (1961), that RT is faster when the stimulus to which reaction has to be made is presented when the EEG indicates physiological alerting as defined by alpha blocking.

This is shown in the present study in several ways. Direct analysis of the RT when the alpha is blocked by the forewarning stimulus (Category I) or when alpha has returned (Category E) or when alpha blocks spontaneously (Category S) showed no significant difference between speed of reaction in these conditions. Even when, as in the case of the imbeciles, there seemed to be a suggestion for such a difference, RTs falling into I and E categories when matched for length of forewarning period were shown to be equal.

More indirectly a difference between the alpha-blocking times between Cond. 1 and 2 has been shown (Fig. 2) for normal and Mongol *S*s. Alpha rhythm remained blocked over longer periods in Cond. 2. This should have resulted, if the EEG activation hypothesis is correct, in an interaction term between intervals and conditions in the analysis of reaction-time scores. There should have been a relationship between RT and foreperiod in Cond. 1, which should not appear, or appear only minimally in Cond. 2. Such an interaction was not found.

In the normal *S*s the findings confirmed that irregularly presented foreperiods produced RTs which did not differ with forewarning length. On the average in Cond. 1, during the .5-, 1-, and 2-sec. foreperiods the alpha remained blocked, but it returned during longer forewarning intervals. It could be suggested that while blocking of alpha rhythm represents "sensory activation" probably not extending beyond 2 sec. this may be superceded by a state of attention directed towards a motor set to respond. Thus one might expect in normals after a return of alpha a mobilized response tendency lasting for fairly long intervals of forewarning. The increase in RT with lengthening foreperiods in subnormals might be due to the absence of such a sustained set. The Mongol *S*s showed a very high percentage of inclusive blocks over a .5-sec. forewarning period. On the other hand, it was shown that they gave very long reaction times over this same warning-stimulus interval. Thus in this group longest reaction times were obtained under conditions of maximal sensory activation as defined by alpha blocking but with possible slow development of motor set.

If on the basis of the present findings it is accepted that the state of the EEG indexes of alpha blocking does not determine RT, some explanation of previous findings must be sought. In the case of Fedio et al. (1961) the comparison is made between RT when there is no alerting signal and when there is a forewarning signal which produces blockade. Thus there is confounding between the effect of the warning and the effect of the alpha blocking.

Lansing et al. (1959) used very short foreperiods up to .3 sec. to enable them to present the stimulus when the alpha had no time to block and compare it with the RT to a slightly longer (.3–1 sec.) foreperiod when alpha had blocked. In contrast, in the present study the "alpha-present" condition was that when the alpha had returned after blocking. It is possible that relatively long RTs occurred in Lansing's experiment in this condition because of motor unpreparedness with a short forewarning.

A further difference between the previous and the present experiments is that they both used an auditory forewarning while the present one used a visual warning signal. Morrell (1961) cites evidence that stimuli of a given sensory modality may produce restricted effects in the cortical projection area for that modality. Thus visual stimuli selectively block the parietooccipital alpha rhythm, while acoustic signals selectively desynchronize the auditory receiving area. While visual stimulation is usually considered the more effective in producing alpha blocking it has been suggested (Bernhaut, Gellhorn, & Rasmussen, 1953) that visual stimulation has a less arousing effect than auditory stimuli. It is therefore possible that the failure of the present findings to relate EEG activity to RT is due to the modality difference between this and previous experiments.

REFERENCES

BERKSON, G., HERMELIN, B., & O'CONNOR, N. Physiological responses of normals and institutionalized mental defectives to repeated stimuli. *J. ment. Defic. Res.*, 1961, **5**, 30–39.

BERNHAUT, M., GELLHORN, E., & RASMUSSEN, A. T. Experimental contributions to the problem of consciousness. *J. Neurophysiol.*, 1953, **16**, 21–35.

FEDIO, P. M., MIRSKY, A. F., SMITH, W. J., & PARRY, D. Reaction time and EEG activation in normal and schizophrenic subjects. *EEG clin. Neurophysiol., Amsterdam*, 1961, **13**, 923–926.

HERMELIN, B. M., & BERKSON, G. Effects of variation in the warning signal on the reaction time of normals and severely subnormals. *Quart. J. exp. Psychol.*, 1964, **16**, in press.

KLEMMER, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51**, 179–184.

LANSING, R. W., SCHWARTZ, E., & LINDSLEY, D. B. Reaction time and EEG activation under alerted and nonalerted conditions. *J. exp. Psychol.*, 1959, **58**, 1–7.

MORRELL, F. Electrophysiological contributions to the neural basis of learning. *Physiol. Rev.*, 1961, **41**, 443–494.

WOODROW, H. The measurement of attention. *Psychol. Monogr.*, 1914, **17**(5, Whole No. 76).

(Received June 3, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 373–378

ROLE OF AWARENESS IN MEANING ESTABLISHED BY CLASSICAL CONDITIONING

BURTON H. COHEN

Lafayette College

The extent to which awareness of the conditioning contingency influenced the classical conditioning of word meaning to nonsense syllables was examined. Using the technique developed by Staats and Staats, nonsense syllables (CS) were paired with words (UCS) of high- and low-evaluative meaning. After the conditioning trials, Ss rated the syllables on a 7-point semantic differential scale whose end points were labeled Pleasant and Unpleasant. 2 experimental groups of 97 Ss and a control group of 34 Ss were tested. Ss were classified as aware or unaware by 3 independent judges who evaluated Ss' written comments about the experiment. Contrary to the results of Staats and Staats, analysis of the data provided no evidence for "learning without awareness."

Following the impetus provided by Greenspoon's (1955) initial study, a large body of literature now exists concerning verbal operant conditioning (e.g., Krasner, 1958). Whereas earlier studies were concerned with the responses to be reinforced and the type of reinforcing stimuli, some recent studies have concentrated on the question of whether conditioning can occur if *S*s are not aware of the conditioning contingency. For example, Dulany (1961), Eriksen (1960), Krieckhaus and Eriksen (1960), Levin (1961), and Spielberger and Levin (1962) have all reached the general conclusion that there is little or no evidence for "learning without awareness." In a similar vein, Paul, Eriksen, and Humphreys (1962) found no evidence for operant conditioning of a human motor response unless *S*s could verbalize the conditioning contingency.

Though relatively few studies have used verbal material as both CS and UCS, it is legitimate to ask whether "learning without awareness" is possible when the classical-conditioning model is used. Staats and Staats (1957) investigated the problem of classical conditioning of word meaning to nonsense syllables. Of 86 *S*s, they found only 9 who could verbalize the correct reinforcement contingency. Their results for "unaware" *S*s clearly showed the predicted conditioning effect.

In view of the recent turn of events with regard to the problem of awareness in operant verbal conditioning, it appeared desirable to re-examine the Staats and Staats (1957) study. Special attention was directed to (*a*) the determination of awareness or unawareness and (*b*) the inclusion of two control conditions not mentioned in the original report.

METHOD

The procedure outlined for Exp. I by Staats and Staats (1957) was followed in all essential details. Initially, *S*s were presented with two tasks designed to train them for the procedure of the actual experiment. The first task involved the visual presentation and recall of a series of nonsense syllables. The second task involved the oral presentation of a series of words. Immediately after *E* spoke each word, *S*s were instructed to repeat the word aloud and then to continue implicitly rehearsing the word until *E* spoke the next word. After all words were presented, *S*s were asked to (*a*) write the words they could recall and (*b*) identify which member of 12 word pairs was a word of the presented list.

The experiment proper consisted of presenting 108 conditioning trials. On each trial one of six nonsense syllables (YOF, XEH, WUH, GIW, LAJ, and QUG) was flashed on a screen by means of a Kodak Cavalcade projector, Model 540. One second after the syllable appeared, *E* called a word aloud. The *S*s were instructed to pay attention to the syllable while at the same time they were to repeat aloud the word spoken by *E* and then implicitly rehearse the word until the next syllable was flashed. Each syllable appeared on the screen for 5 sec. Each of the syllables appeared 18 times and was paired with a different word each time. The syllable-word pairs were the same as those reported in Table 1 of Staats and Staats (1957) and were presented in the order given in the table. As determined from semantic differential scaling, 18 words had high-evaluative meaning and 18 words had low-evaluative meaning. The remaining 72 words had no systematic meaning.

Immediately after the conditioning trials, *S*s were asked to check the position of each of the syllables on a 7-point semantic differential scale whose end points were labeled Pleasant and Unpleasant. Along with the six original syllables, three additional syllables (MEQ, FEH, and SIJ) were included as a control measure. All nine syllables appeared, one per page, in a booklet distributed to *S*s. A large variety of random orderings of the pages of the booklets was obtained during the collating process.

After *S*s had scaled the syllables, they were requested to write on the back of their booklet "anything you thought about the experiment, especially what you think the purpose of the experiment was, and anything you thought about during the experiment."

In all, six subgroups of male, undergraduate Lafayette College students served as *S*s. The groups varied in size from 13 to 18 *S*s. Half the subgroups ($N = 47$) had YOF paired with the high-evaluative words and XEH paired with the low-evaluative words and will be designated as Group I. For the other half of the subgroups ($N = 50$), designated Group II, the pairings were reversed.

To determine the mean scale values of the syllables in the absence of any conditioning trials, an independent group of 34 male, undergraduate Lafayette College students served as *S*s in a control condition. These *S*s were presented with the 9-page booklets and given the same instructions as experimental *S*s with regard to the semantic differential scaling of the syllables.

For more explicit details about the experimental procedure, the reader is referred to the original article (Staats & Staats, 1957). The three major changes from the procedure of Staats and Staats were as follows. (*a*) The experimental *S*s scaled nine rather than six syllables. (*b*) The ordering of the syllables in the booklets was randomized. (*c*) A control group of naive *S*s scaled all nine syllables.

RESULTS

The *S*s' written comments about the experiment were evaluated by three independent judges. The judges were familiar with the design of the experiment and were asked to indicate whether *S*s' comments indicated awareness or unawareness of the conditioning contingency. For the 97 experimental *S*s, the results of this judgmental operation were as follows. All three judges agreed that 45 *S*s showed no awareness and that 29 *S*s indicated awareness of the contingency. Of the remaining 23 *S*s, two judges agreed that 14 *S*s were aware and one judge indicated that 9 *S*s were aware. For purposes of data analysis, *S*s were regarded as being aware of the contingency if at least one judge rated their comments as such. On this basis, 52 *S*s were classified as aware. Thus, the classification of *S*s as unaware was quite stringent while the classification of *S*s as aware was somewhat conservative. Table 1 presents the mean scale values for each of the nine syllables for the aware and unaware *S*s of the two experimental groups and for the control group. Also included in Table 1 are the combined means for various groupings of the syllables.

Several features of the data in Table 1 should be noted immediately. (*a*) With regard to the data of control *S*s, a considerable difference in the mean scale values of the experimental

TABLE 1

INDIVIDUAL AND COMBINED MEAN SCALE VALUES FOR THE NONSENSE SYLLABLES

Group		Individual *M*s									Combined *M*s			
		(1) YOF	(2) XEH	(3) GIW	(4) LAJ	(5) QUG	(6) WUH	(7) FEH	(8) MEQ	(9) SIJ	(1–6)	(3–6)	(7–9)	(1–9)
I	Aware $N = 24$	2.38	5.21	3.54	2.75	3.50	3.63	4.38	4.29	4.67	3.50	3.35	4.44	3.81
	Unaware $N = 23$	3.30	4.30	3.74	2.70	4.04	4.00	4.57	4.48	3.57	3.68	3.62	4.20	3.86
II	Aware $N = 28$	5.57	1.82	3.61	3.07	4.18	3.71	3.61	3.86	4.14	3.66	3.64	3.87	3.73
	Unaware $N = 22$	4.05	4.55	2.91	2.55	4.77	3.91	3.68	4.32	4.05	3.79	3.53	4.02	3.86
	Control $N = 34$	3.71	4.82	3.82	3.03	5.15	3.74	4.56	4.12	3.82	4.04	3.93	4.17	4.08

Note.—The Pleasant and Unpleasant ends of the semantic differential scale were represented by scale values of 1 and 7, respectively. Even though the distributions of scale values for some syllables were far from normally distributed, mean rather than median scale values are reported because of the discrete nature and restricted range of the data.

syllables (YOF and XEH) was apparent in the absence of any prior conditioning. As might be expected on the basis of pronunciability (Underwood & Schulz, 1960), YOF was rated more pleasant than XEH. Thus, a valid assessment of the difference in mean scale values for YOF and XEH for the experimental groups should be evaluated against this control discrepancy. (*b*) For all groups, there was considerable variability in the mean scale values for the syllables. (*c*) Inspection of the combined means shown in Table 1 revealed that there was a general tendency to rate all six syllables used in the conditioning procedure towards the pleasant end of the scale. The control data showed no such tendency. This displacement of mean scale values is in general agreement with a large body of literature which has demonstrated that pleasant words, events, experiences, etc., are retained better than unpleasant words. On the other hand, the combined means for the syllables which were not used in the conditioning procedure were essentially similar for the experimental and control groups.

The major analyses of the results concerned the comparison of scale values for YOF and XEH. It should be pointed out that use of a response measure based on a 7-point scale seriously violates the assumptions underlying the use of parametric tests of significance since (*a*) the scale values were discrete and quite restricted in range and (*b*) the distribution of scale values for a number of the syllables (especially in the case of YOF and XEH for aware *S*s) departed considerably from any approximation to a normal distribution. However, as a first stage in the data analysis, the overall means for YOF and XEH for Groups I and II (i.e., disregarding the aware-unaware classification) were analyzed by the same analysis of variance technique used by Staats and Staats. For comparison purposes, the overall means for the present study are contrasted with the corresponding means

TABLE 2

Comparison of Staats and Staats Results with Those of Overall Mean Scale Values of Present Study

Group	Study			
	Staats and Staats		Present	
	YOF	XEH	YOF	XEH
I	2.40	4.80	2.83	4.77
II	4.73	3.13	4.90	3.02

of the Staats and Staats study and are shown in Table 2. The general agreement between the two studies is quite striking (even though Staats and Staats maintain that aware *S*s were eliminated from their data). The results of the analysis of variance of the data from the present study were in close agreement with the Staats and Staats analysis. The only significant effect was that of Conditioning of Meaning, $F(1, 95) = 30.57, p < .001$. In contrast, as shown in Table 3, no significant *F*s were obtained from the analysis of variance of data for unaware *S*s of Groups I and II.

Using nonparametric tests, a number of other comparisons were made. The Wilcoxen matched-pairs, signed-rank test was used to evaluate the mean difference between YOF and XEH for the control group. Using the normal approximation, the resulting value ($z = 2.35, p < .02$) confirmed the notion that YOF received a significantly lower mean scale value than XEH. The difference in mean scale values between XEH and YOF ($4.82 - 3.71 = 1.11$) provided a baseline against which to evaluate the mean difference between these syllables for *S*s of Groups I and II. For unaware *S*s of Groups I and II, the differences were 1.00 and 0.50, respectively. For aware *S*s of Groups I and II, the differences were 2.83 and −3.75, respectively. Not only did the differences for unaware *S*s fail to exceed the control difference, but for Group II the difference was in the wrong direction.

A further set of evaluations compared the cumulative distribution of scale values of YOF and XEH for *S*s of Groups I and II with the corresponding cumulative distributions of scale values for the control group. The conservative χ^2 approximation for the two-sample Kolmogorov-Smirnov test was used (see Goodman, 1954). The results are shown in Table 4. In no case did the comparisons of the distributions for unaware *S*s with the control data produce anything close to a significant result.

Figure 1 presents a graphic view of

TABLE 3

Analysis of Variance of Conditioning Data for Unaware *S*s

Source	*df*	*MS*	*F*
Between *S*s			
Groups	1	5.42	1.48
Error	43	3.66	
Within			
Syllables	1	12.84	3.71
Cond. Meaning	1	1.41	<1.0
Residual	43	3.46	

TABLE 4

Comparison of Distribution of Scale Values of YOF and XEH by Kolmogorov-Smirnov Test

	Comparison	χ^2
Aware	Group I (YOF) vs. Control (YOF)	12.17**
	Group I (XEH) vs. Control (XEH)	9.32**
	Group II (YOF) vs. Control (YOF)	12.32**
	Group II (XEH) vs. Control (XEH)	32.99**
Unaware	Group I (YOF) vs. Control (YOF)	1.09
	Group I (XEH) vs. Control (XEH)	1.15
	Group II (YOF) vs. Control (YOF)	0.77
	Group II (XEH) vs. Control (XEH)	2.03

** $p < .01$.

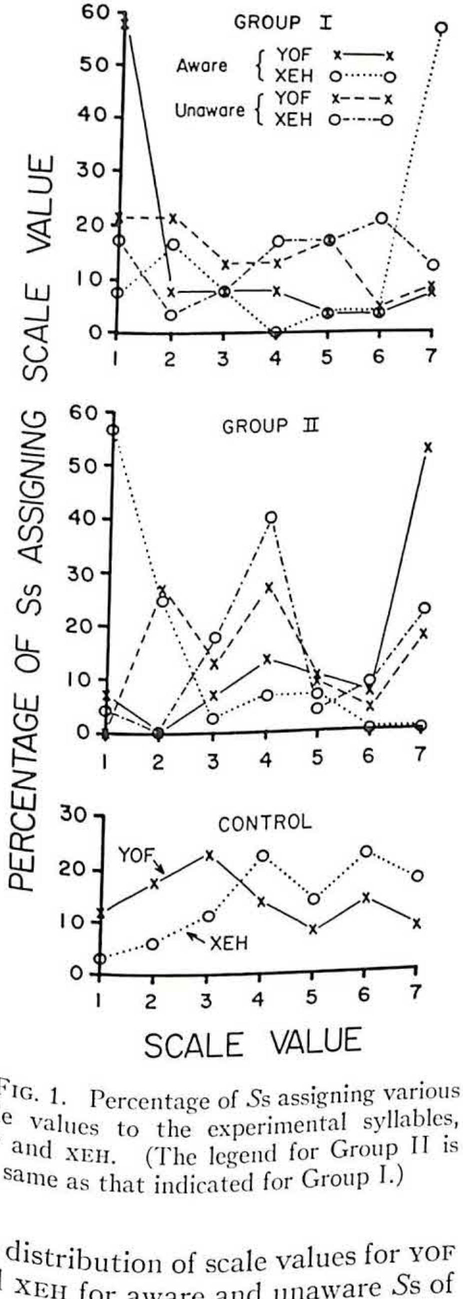


FIG. 1. Percentage of *S*s assigning various scale values to the experimental syllables, YOF and XEH. (The legend for Group II is the same as that indicated for Group I.)

the distribution of scale values for YOF and XEH for aware and unaware *S*s of Groups I and II as well as the baseline data of the control group. Inspection of Fig. 1 provides striking evidence of the lack of any conditioning effect for unaware *S*s.

DISCUSSION

The results of the present study seem unequivocal: *S*s who were aware of (i.e., could verbalize) the conditioning contingency showed the predicted effect while *S*s who were unaware did not show the effect. Why then the discrepancy between the present results and those of Staats and Staats?

In reconstructing the detailed experimental procedure used by Staats and Staats, it was possible that inadvertent cues were given to *S*s of the present study which had the tendency of increasing the extent of awareness. Subsequent information about the experimental procedure[1] gave no support for this notion. Even if such were the case, the fact remains that *S*s, classified as unaware, showed no conditioning effect. It should also be pointed out that though *S*s of this experiment were all enrolled in an introductory psychology course, the experiment was conducted during the first week of the semester. Thus, it seems unlikely that "sophistication" regarding the scope or design of the experiment could have influenced the results.

The close agreement between the overall results of this study and those of Staats and Staats (Table 2) suggests the possibility that all aware *S*s were not eliminated from the data reported by Staats and Staats. However, since Staats and Staats also examined *S*s' written comments, it seems unlikely that they failed to eliminate any *S*s whose comments clearly delimited the conditioning contingency. The most plausible explanation seems to hinge on the evaluation of the written comments of *S*s. It is possible that *S*s of the present experiment wrote longer and more detailed protocols than *S*s of the Staats and Staats study. This might have allowed for the elimination of more aware *S*s than would be possible if short, terse comments were written. In any event, within the framework of a verbal classical-conditioning model, this study provides no support for "learning without awareness." Further verification of this finding would have wide-spread implications for current theories which where $\Delta\alpha$ is a small amount of binocular disparity, $\Delta d'$ is the perceived depth interval which results from $\Delta\alpha$, and S' is a perceived size of a frontoparallel extent whose retinal size (visual angle) is θ. The term C is a constant for a particular *O*. The ratio S'/θ is the perceived size per unit of retinal size at the distance of the objects producing $\Delta\alpha$. As a ratio, S'/θ is independent of any particular frontoparallel extent and can be applied to an entire frontoparallel plane.

---

[1] A. W. Staats, personal communication, 1963.

It is evident that S'/θ usually is not constant throughout the visual field. For example, if perfect frontoparallel size constancy is present, for a constant value of θ, S' and hence S'/θ will increase linearly with the physical distance from *O*. It is clear, therefore, that according to Equation 1, a constant value of binocular disparity will produce different perceived extents when it is perceived to be in portions of the visual field having different values of S'/θ. It is the purpose of the present experiment to test this point of view. Objects producing a constant value of binocular disparity will be made to appear in portions of a field-of-view having different values of S'/θ. The resulting perceived depth interval should increase as the value of S'/θ increases.

METHOD

Apparatus

Binocular stimuli.—A perspective schematic drawing of the apparatus is given in Fig. 1. A binocular disparity was generated by binocularly viewing a configuration produced by two pairs of objects. One pair consisted of two luminous half-circles (Objects *e* and *f*) and the other pair consisted of two luminous rectangles (Objects *g* and *h*). The four objects (*e*, *f*, *g*, and *h*) were physically located to the right of *O* (see Fig. 1). Light from these objects passed through a black tunnel and was reflected from a mirror into the eyes of *O*. The result was that *O* perceived the objects to be in front of himself instead of at their actual locations. The perceived positions of Objects *e*, *f*, *g*, and *h* are given in Fig. 1 as *e′*, *f′*, *g′*, and *h′*, respectively.

As is illustrated in Fig. 1, Objects *e*, *f*, *g*, and *h* were formed by placing a diffused fluorescent source of light (94 ft-L) behind appropriate apertures (Lichten & Lurie, 1950). Care was taken that no light from the portion of the apparatus producing Objects *e*, *f*, *g*, and *h* was visible except that from the objects. Both Objects *g* and *h* (the luminous rectangles) were physically 277.2 cm. from *O* while both Objects *e* and *f* (the luminous half-circles) were physically 252 cm. from *O*. The half-circles were always higher in the visual field (by 30′ of visual angle) than the luminous rectangles. For the average interpupillary distance of the *O*s used in this study (6.64 cm.), the binocular disparity between the pair of luminous rectangles and the pair of half-circles was 477″ of arc. The height and width of Objects *g* and *h* subtended visual angles of 45′ and 22.5′ of arc, respectively. The size of Objects *e* and *f* were such as to subtend a visual area equal to that subtended by Objects *g* and *h*. The horizontal separation between Objects *e* and *f* or between Objects *g* and *h* was 2°57′ of visual angle.

The physical position of Objects *e*, *f*, *g*, and *h* was constant throughout the study. Thus, both the convergence values of the eyes to Objects *e*, *f*, *g*, and *h* in the binocular configuration and the binocular disparities between these objects were invariant throughout this study.

The alley.—As illustrated by Fig. 1, Objects *e′*, *f′*, *g′*, and *h′* appeared to be located in an alley which extended away from *O* in distance. The floor of the alley (48 ft. long and 4 ft. wide) consisted of a checkerboard pattern of light and dark gray rectangles (8 in. wide and 16 in. long). The alley was illuminated by a series of overhead fluorescent lights extending the length of the alley. The sides of the alley were formed on the right by white curtains and on the left by a wall covered with white cloth. The alley and all objects in the alley except Objects *e*, *f*, *g*, and *h* were observed monocularly. This was accomplished by opaquing the back surface of that portion of the mirror which was in front of the left eye of *O*. The mirror was a first-surface mirror which both reflected and transmitted light. This enabled *O* to view Objects *e*, *f*, *g*, and *h* with both eyes while viewing the alley with his right eye only. Care was taken that the left eye of *O* was in darkness except for Objects *e*, *f*, *g*, and *h*.

To remove extraneous light, the viewing position of *O* was enclosed in black velveteen and a small black tube extended from the left eye of *O* to the vicinity of the left half of the

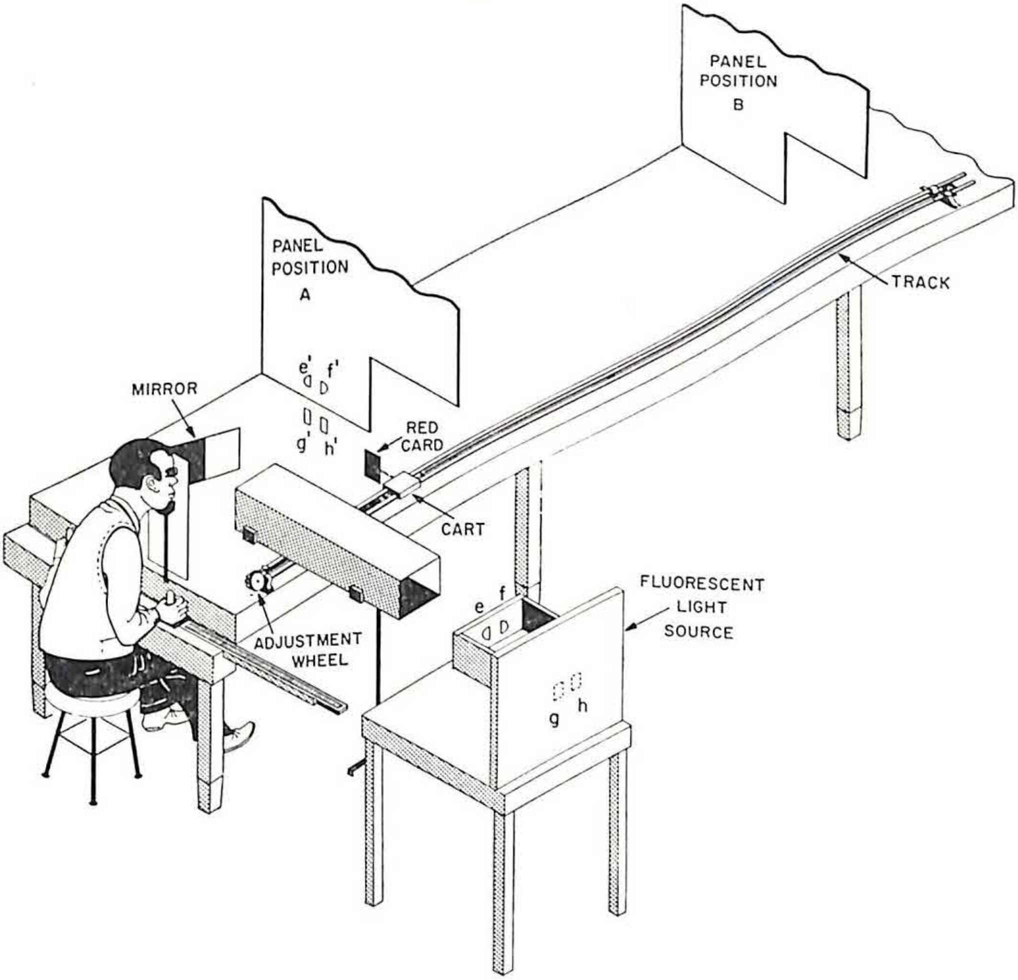


FIG. 1. A schematic perspective drawing of the apparatus used in the study.

mirror. To eliminate the accommodative cue between objects in the visual field, *O* viewed the entire scene through two pinhole apertures (.05 cm. in diameter) located very close to his eyes. When a small black curtain in front of the viewing position was lowered, the view to the monocularly observed alley was cut off without disturbing the ability to see the binocular stimuli. An adjustable chin rest was located at the viewing position of *O* and a light-adaptation surface (51 ft-L), which could be turned on or off by *E*, was located to the left of the viewing position.

To make the constant binocular configuration appear in different portions of the monocularly observed alley at different times, a monocularly observed panel consisting of diagonal black and gray stripes was vertically presented at either 182.9 or 548.6 cm. from *O* in the alley. The two alternate positions of the panel are illustrated as A and B of Fig. 1. It was anticipated that the binocular configuration would appear to *O* to be closer to himself (in a less distant portion of the monocularly observed alley) when the panel was in Position A rather than in Position B. Since some amount of frontoparallel size constancy would be expected to occur in the alley, the different distance positions in the alley would be expected to have different values of S'/θ. The process by which the panel would make the binocular configuration appear to change in distance position in the monocularly observed alley is called the "equidistance tendency" (Gogel, 1956). This is the tendency for objects to appear equidistant unless restricted by the presence of effective depth cues. It is clear that, except for the equidistance tendency, there was no reason why the binocular configuration would appear at one distance rather than another in the alley. The only factor other than the equidistance tendency available for localizing the binocular configuration in apparent distance was the convergence of the eyes required for binocular fusion. But it has been demonstrated that this cue is usually ineffective in establishing a perceived distance (Gogel, 1962b). Thus, the equidistance tendency would be expected to be effective between the binocularly observed objects and the monocularly observed panel with the result that the binocular configuration would appear to be at the distance vicinity of the panel. Other monocularly observed objects in the alley also would be expected to have an "equidistance" effect upon the binocularly observed objects. However, since the strength of the equidistance tendency is inversely related to directional separation (Gogel, 1956), the most important determiner of the apparent position of the binocular objects would be the result of the equidistance tendency with respect to the panel.

Measurements of apparent size and distance.—Three types of psychological measurements were required in this study. It was necessary to determine (*a*) that the change in the position of the panel was effective in inducing a change in the apparent position of the binocular configuration (Objects *e′*, *f′*, *g′*, and *h′*), (*b*) that concomitant with the change in apparent position of the binocular configuration, the value of S'/θ associated with the frontoparallel extents of the binocular configuration also changed, and (*c*) that the apparent depth within the binocular configuration changed in agreement with the change in S'/θ.

The measurement of the apparent depth position of Objects *e*, *f*, *g*, and *h* was accomplished by using the track and cart (see Fig. 1). Attached to the cart by means of a black rod was a red card (10.1 cm. wide × 12.7 cm. high) presented vertically on the floor of the alley. The bottom of the red card always made contact with the floor of the alley. The red card was monocularly observed (with the right eye) and its distance position in the alley was clearly seen by the interposition of its bottom edge with the alley floor. By turning a knob with his right hand, *O* could move the cart along the track and consequently could move the red card in distance along the alley to the apparent depth position of Objects *e*, *f*, *g*, or *h* as required.

To measure the values of S' in the binocular configuration, a distance indicator using the kinesthetic modality was located at the position of *O*. The device consisted of two vertical rods which *O* could grasp, one in each hand. The left rod was stationary while the right rod could be moved laterally until the separation between the rods seemed to *O* to be the same as a designated frontoparallel distance (width) in the binocular configuration. The hand-adjustment apparatus was not visible to *O* during the width adjustments. An attached scale permitted *E* to measure the lateral distance (S') which *O* adjusted between the rods. A similar device has been found in previous studies to be a valid measure of perceived size (Gogel, 1962a; Gogel, Wist, & Harker, 1963). From the visual angle (θ) of the width being perceptually duplicated and the obtained adjustment (S'), the ratio S'/θ could be computed.

The apparent depth within the binocular configuration (the apparent depth from the rectangles to the half-circles) was also measured by means of the hand-adjustment apparatus. The *O*s kinesthetically adjusted the lateral distance between the rods until it appeared equal to the designated depth interval in the binocular configuration.

Observers

Ten men and six women were used as *O*s. Each *O* had a stereoscopic acuity of at least 18″ of arc as measured with the Keystone Multi-Stereo Test. As a consequence of a preliminary test, several potential *O*s were not used in the experiment because they saw some portion of the binocular configuration imbedded in the surface of the panel.

Procedure

To determine that *O* saw the alley as extending away from himself in distance, a white (monocularly observed) target, 7.6 cm. wide and 20.3 cm. high, was presented vertically on the floor of the alley at randomly presented distances of either 140, 235, 330, 425, or 520 cm. from *O*. Each *O* adjusted the red card to apparent equidistance with the white target at each of the physical distances, and also made verbal estimates in feet of the distance of the white target from himself. During this procedure, the striped panel was located at its far position. Following these adjustments, the white target was removed and the front curtain was lowered, eliminating the view of the alley. The binocular stimuli were turned on and *O* checked to determine that he saw all four objects of the binocular stimulus (*e*, *f*, *g*, and *h*) with each eye. The front curtain was raised and *O* was asked to adjust the lateral distance between the inner edges of the rods of the hand-adjustment apparatus until it seemed to be the same physical size as the extent between either: (*a*) the inner edges of the half-circles (a judgment of width, S'_{ef}) or; (*b*) the inner edges of the rectangles (a judgment of width, S'_{gh}) or; (*c*) the depth between the left half-circle and the left rectangle (a judgment of depth, d'_{eg}) or; (*d*) the depth between the right half-circle and the right rectangle (a judgment of depth, d'_{fh}). For a particular position of the panel and a particular *O*, all judgments of width (or depth) were completed before proceeding to the judgments of depth (or width).

Thus, each *O* used the hand-adjustment apparatus to make two width and two depth judgments for each of the two positions of the panel. Each *O* repeated each width or depth judgment three times. The starting position for the hands for these adjustments (hands together or hands far apart) was systematically varied. Following each set of three width or depth judgments, *O* adjusted the red card to the distance of each of the two binocular objects involved in the particular judgment, verbally estimated in feet the distance to each binocular object, and turned to look at the light-adaptation surface which was turned on by *E* for approximately 10 sec. After completing the width and depth judgments for one position of the striped panel, *O* adjusted the red card to the distance of the panel, and verbally estimated this distance from himself in feet. Following this, the striped panel was moved to the second distance and the judgments were repeated. With half the *O*s, the panel was presented first at the near position and then at the far position. With the remaining *O*s, the order of presentation was reversed.

RESULTS

Figure 2 contains the average results expressed in centimeters obtained both from the adjustment of the red card to the apparent distances of the white target and from the verbal reports of the distances of the white target. It is clear from Fig. 2 that the alley appeared to extend away from *O* in distance and that increasing values of cart adjustment (or verbal reports) represent increasing values of perceived distance.

The mean and median results in centimeters for the hand adjustments are given in Table 1. The measurements of perceived depth between the left and right pairs of binocular objects have been averaged in the *d′* data of Table 1. For two *O*s, the perceived depth within the binocular configuration was greater than could be measured with the hand-adjustment apparatus. These two *O*s were asked to perceptually duplicate only $\frac{1}{3}$ of the magnitude of each extent within the binocular configuration and, therefore, the results from these *O*s were multiplied by 3 to arrive at the values which would be expected to occur if these *O*s had been able to make the apparent

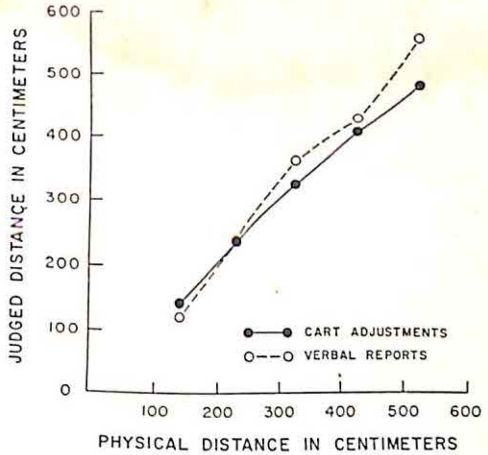


FIG. 2. The relation between behavioral and physical distance in the alley.

TABLE 1

DEPTH AND WIDTH ADJUSTMENTS IN CENTIMETERS AS A FUNCTION OF THE APPARENT POSITION OF THE CONFIGURATION

Distance Position of Panel (Cm.)	Depth Adjustments (d')		Width Adjustments (S')			
			Between Half-Circles (S'_{ef})		Between Rectangles (S'_{gh})	
	M	Mdn.	*M*	Mdn.	*M*	Mdn.
548.6	68.2	42.1	38.4	37.0	45.3	37.0
182.9	37.8	26.0	31.5	31.3	34.8	23.2
Difference	30.4	16.1	6.9	5.7	10.5	13.8

unit of visual angle of frontoparallel objects (S'/θ), not the change in perceived absolute distance, was the significant variable.

The viewpoint expressed in the present study illustrates the importance of so-called "monocular cues" upon the depth perception resulting from binocular cues. Many cue systems affect S' and, through S'/θ, affect the perception of depth from binocular disparity. Thus, the monocular cues, to the extent that they determine S', also determine the perceived depth associated with a particular binocular disparity. This statement is not equivalent to the truism that a variety of cue systems contributes to the perception of depth. The assertion is that the perceived depth per unit of binocular disparity ($\Delta d'/\Delta\alpha$ of Equation 1) is not predictable unless the perceived size per unit of visual angle of frontoparallel extents (S'/θ) in the depth vicinity of the disparity is known. Except in the case of zero amount of size constancy, the value of S'/θ will differ for different distances in the visual field. Unless the values of S'/θ are known, the perceived depth from a binocular disparity in any specified part of the visual field cannot be predicted. The amount of size constancy present and the magnitude of the perceived depth between binocular objects are interrelated phenomena and a theory of binocular vision is incomplete unless it provides for this interrelation.

Visual judgments of extent occur also in situations in which only binocular cues to size and distance are available. It is asserted that in these situations also, the perceived depth per unit of binocular disparity ($\Delta d'/\Delta\alpha$) and the perceived size per unit of visual angle of frontoparallel extents (S'/θ) will vary concomitantly as a function of distance. According to Equation 1, $\Delta d'/\Delta\alpha$ and S'/θ are proportional in any situation involving depth intervals determined by binocular disparity cues.

REFERENCES

Fry, G. A. Visual perception of space. *Amer. J. Optom.*, 1950, **27**, 531–553.

Gogel, W. C. The tendency to see objects as equidistant and its inverse relation to lateral separation. *Psychol. Monogr.*, 1956, 70(4, Whole No. 411).

Gogel, W. C. The effect of convergence and angular size upon the computed value of a hypothetical observer constant in binocular vision. *USA Med. Res. Lab. Rep.*, 1958, No. 372. (a)

Gogel, W. C. An observer constant in the perception of stereoscopic depth. *USA Med. Res. Lab. Rep.*, 1958, No. 316. (b)

Gogel, W. C. Perceived frontal size as a determiner of perceived binocular depth. *J. Psychol.*, 1960, **50**, 119–131. (a)

Gogel, W. C. The perception of a depth interval with binocular disparity cues. *J. Psychol.*, 1960, **50**, 257–269. (b)

Gogel, W. C. The perception of shape from binocular disparity cues. *J. Psychol.*, 1960, **50**, 179–192. (c)

Gogel, W. C. Convergence as a determiner of perceived absolute size. *J. Psychol.*, 1962, **53**, 91–104. (a)

Gogel, W. C. The effect of convergence on perceived size and distance. *J. Psychol.*, 1962, **53**, 475–489. (b)

Gogel, W. C. The visual perception of size and distance. *Vis. Res.*, 1963, **3**, 101–120.

Gogel, W. C., Wist, E. R., & Harker, G. S. A test of the invariance of the ratio of perceived size to perceived distance. *Amer. J. Psychol.*, 1963, **76**, 537–553.

Lichten, W., & Lurie, S. A new technique for the study of perceived size. *Amer. J. Psychol.*, 1950, **63**, 280–282.

Ogle, K. N. Precision and validity of stereoscopic depth perception from double images. *J. Opt. Soc. Amer.*, 1953, **43**, 906–913.

Ogle, K. N. Theory of stereoscopic vision. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill, 1959. Pp. 362–394.

Von Kries, J. Notes on perception of depth. In H. V. Helmholtz, *Treatise on physiological optics.* Vol. 3. (Trans. and ed. by J. P. C. Southall) Menasha, Wisc.: Banta, 1925. Pp. 369–400.

(Received June 10, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 387–391

BACKWARD CONDITIONING AND UCR LATENCY

MILTON A. TRAPOLD

University of Minnesota

MARVIN HOMZIE AND EDWARD RUTLEDGE

University of Iowa

According to Jones' 2-process theory of classical aversive conditioning, backward presentation of the CS and UCS at a temporal separation exactly equal to the UCR latency should lead to the fastest conditioning on the early trials. This prediction was tested by attempting simultaneous backward conditioning of the long-latency GSR and the short-latency eyeblink and finger-withdrawal responses, using a 1,500-msec. UCS-CS interval in one group and a 40-msec. interval in the other. Contrary to prediction, there was no evidence for conditioning of the short-latency responses under either UCS-CS interval, but in line with prediction, the 1,500-msec. group showed larger conditioned GSRs after only 2 UCS-CS pairings and then remained superior to the 50-msec. group for the rest of the 30-trial conditioning session.

In a recent theoretical article, Jones (1962) has elaborated a theory of classical aversive conditioning which attempts to integrate the data from numerous studies of conditioning as a function of CS-UCS interval. Briefly, this theory views conditioning as involving two conceptually independent processes, both of which contribute to the observed changes in response probability that constitute conditioning.

The first of these two processes (the contiguity process) is a positive increasing function of the temporal proximity of the CS and the UCR. It is assumed that the closer in time (irrespective of order) the CS and UCR, the greater is the increment per trial in the capacity of the CS to evoke a CR.

The second process (the reinforcement process) depends upon the contiguity process already having caused some CRs to occur and is assumed to be a positive increasing function of the temporal proximity of the CR and UCS. Thus, once some CRs have begun to occur, the closer in time the CR is to the UCS (again, irrespective of order), the greater will be the increment per trial in the capacity of the CS to evoke a CR on subsequent trials.

Observed changes in response probability, then, are assumed to be determined by a summation of these two processes.

As Jones has clearly shown, this kind of a theory makes a number of novel predictions about numerous aspects of conditioning as a function of CS-UCS interval and UCR latency.

Of particular interest for the present paper are the predictions about the effects of backward-conditioning procedures. In particular, this theory predicts that on the early trials of conditioning, when the contiguity process is providing the predominant part of any observed changes in response probability, a backward presentation of the CS and UCS should lead to the fastest rise in the conditioning curve. Since the UCR always has some finite latency following the onset of the UCS, the closest temporal proximity of the CS and UCR would be achieved when the UCS precedes the CS by exactly the UCR latency. In this way, the CS and UCR would be made to occur simultaneously, and thus we would expect the fastest conditioning on the basis of the contiguity process alone.

It should be noted that this prediction only holds for the early period of conditioning when the contiguity process is clearly dominant. Once some CRs have begun to occur, the reinforcement process will also come into play and alter this relationship.

In a recent experiment, Champion and Jones (1961) have provided some support for this prediction by showing a slightly greater increase in GSR amplitude following a single backward-conditioning trial with a 750-msec. interstimulus interval than following a single forward-conditioning trial with a 500-msec. interstimulus interval. Following this first trial, however, the forward group rose to a significantly higher level of responding than the backward group and remained there throughout the rest of conditioning. Similarly, Champion (1962) has shown a faster initial rise in the GSR conditioning curve when the UCS precedes the CS by 1,200 msec. than when it precedes it by 2,800 msec. or follows it by 400 msec.

A related implication of this theory is that the optimum backward-interstimulus interval for conditioning during this early period will be related to the UCR latency: the longer the UCR latency, the more the UCS must precede the CS if the UCR and CS are to occur simultaneously.

The purpose of the present study is to collect data on these implications of Jones' theory by attempting backward conditioning of three responses simultaneously. Two of these responses (eyeblink and finger withdrawal) have a relatively short UCR latency while the third (GSR) has a relatively long UCR latency. Thus, according to Jones' theory, one would expect the short-latency responses to show the faster increase on early trials when the UCS precedes the CS by a short time interval and the long-latency responses to show the faster initial increase with a relatively long UCS-CS separation.

METHOD

Subjects.—The *S*s were 14 male and 15 female volunteers from an elementary course in psychology.

Apparatus.—Throughout the experiment, *S*s were seated in an adjustable chair in a semisoundproof room. The recording equipment and stimulus controls were housed in a separate room from which *E* could communicate with *S* via an intercom system.

The apparatus for recording the eyelid responses and presenting the CS and air puff was identical to that used in previous eyelid conditioning experiments from the Iowa laboratory (cf. Spence, 1953).

The CS was a 500-msec. increase in brightness (.004 ml.–3.00 ml.) of a 2.25-in. diameter circular milk-glass disk, located approximately 4 ft. directly in front of *S*. The UCS was a combination of a 2-psi air puff of 50 msec. duration delivered through a .062-in. diameter oriface to the right cornea and an electric shock delivered to the middle finger of the left hand. The shock was provided by the discharge of a 4-μf. capacitor, charged by a variable-voltage power supply, through copper electrodes 1 in. long and .25 in. wide. One of these electrodes was taped to either lateral surface of the middle finger of the left hand, distal to the third joint. Electrode paste was used to provide good skin contact over the entire area of the electrodes.

The GSR was recorded by means of a 1-in. diameter zinc electrode strapped to each of *S*'s palms, a Hunter GSR amplifier, and an Esterline-Angus recording milliammeter. A 5% zinc sulphate electrode paste was used on the electrodes to insure minimal electrode contact resistance.

Finger-withdrawal responses were recorded by means of a small copper disc taped to but insulated from the tip of the middle finger of *S*'s left hand in such a way that when the hand was relaxed, the disc rested on a metal plate on the chair arm. This contact closed a circuit to a marker pen of the polygraph used to record eyelid responses. Thus, when *S*'s finger jumped off the arm of the chair, this contact was broken and the marker pen deflected.

Procedure.—When *S* arrived at the experimental room, he was first asked to wash his hands and then was seated in the chair and the various pieces of apparatus were connected to him. These consisted of (*a*) the headset containing the eyelid-movement sensor and the air-puff nozzle; (*b*) the two GSR electrodes; (*c*) the two shock electrodes; and (*d*) the metal disc for recording finger withdrawals.

After these were connected and the apparatus was ascertained to be functioning properly, *S* was read a prepared set of instructions which described the experiment as being concerned with collecting normative data on certain muscular and physiological responses to a number of stimuli delivered in various combinations. The *S* was cautioned that he would receive air puffs, shocks, and lights, that he should look at the light at all times, and that the shocks would cause his finger to jump slightly off the arm of the chair. The *S* was requested not to attempt to restrict this movement of his finger but to be sure to replace his finger on the metal plate as soon as the shock had subsided.

Following these instructions, the shock level to be used during the experiment proper was determined for each *S*. This was accomplished by giving *S* a series of shocks starting at 30 v. and increasing by 10 v. on each successive trial. The *S* was instructed to tell *E* when the shock became as intense as he thought he could stand. This, then, was the shock level that was used throughout the experiment. If *S* did not stop the shock increase by 150 v., no further increases were made, and the 150-v. level was used on all subsequent presentations. Regardless of what shock intensity *S* indicated as being his limit, however, further shocks were given until *S* had received a total of 15 shock presentations during this preliminary phase. On the last three of these shock presentations, the air puff was simultaneously presented.

Following this preliminary procedure, *S* also received three presentations of the CS alone. These trials provided a measure of the capacity of the CS to evoke the various responses prior to actual pairing of the UCS and CS, but following a series of shock and air puffs similar to that which would precede test trials during actual conditioning.

The experiment proper was then begun. This consisted of a series of 30 backward-conditioning trials on which the UCS was presented followed by the CS and interspersed among which were seven test trials on which the CS alone was presented. The sequence of conditioning and test trials was the same for all *S*s, the test trials being presented after Conditioning Trials 2, 5, 9, 14, 20, 25, and 30. Group 40 had an interval of 40 msec. separating the onset of the UCS and the onset of the CS (the approximate average latency of the unconditioned eyelid reflex with a 2-psi air puff), and Group 1,500 had an interval of 1,500 msec. separating the onset of the UCS and the CS (the approximate average latency of the unconditioned GSR response under these conditions of evocation). Group 40 was comprised of eight males and seven females and Group 1,500 of seven males and seven females.

The test trials were administered exactly as the conditioning trials except for the omission of both UCSs. The intertrial interval separating all trials was determined to a large extent by how long the GSR took to return to a stable level after each trial, but 45 sec. was adopted as an arbitrary minimum intertrial interval.

RESULTS

The results of this experiment are summarized in Fig. 1 where the performance of both groups on all three responses is plotted as a function of the number of conditioning trials. A number of points are rather immediately obvious from this figure. First, finger withdrawals to the CS were extremely rare. Of a total of 203 trials (29 *S*s × 7 test trials) on which such a response could have been observed, only 7 responses were actually recorded, 5 in the 1,500-msec. group and 2 in the 40-msec. group. This fact, taken in combination with the curves in Fig. 1, makes any attempt to determine a statistical difference between the two groups superfluous.

Second, while the absolute frequency of eyeblinks was somewhat larger than the finger withdrawals, there was again no systematic difference between the overall response

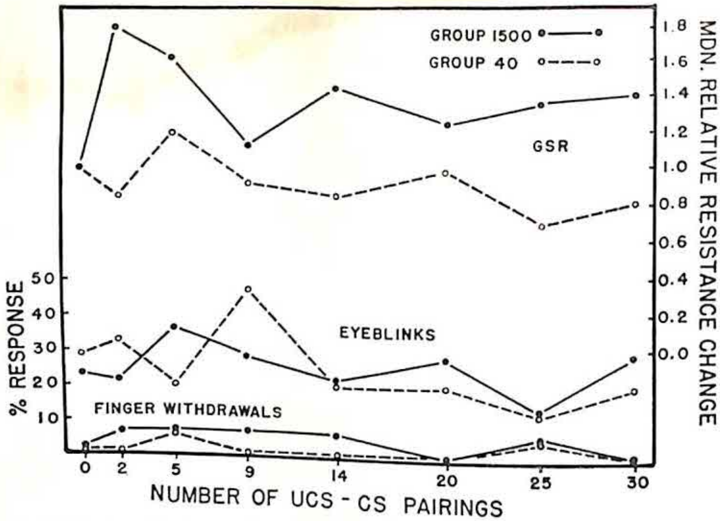


FIG. 1. Backward-conditioning curves for GSR, eyeblink, and finger-withdrawal responses.

levels or the trends of the two groups. The total number of responses over all test trials was exactly 25 for both groups.

For the GSR response, a somewhat different picture obtains. A rather large difference in favor of the 1,500-msec. group was present on the first test trial and was maintained throughout the remainder of conditioning. A Mann-Whitney *U* test of the difference between the two groups on relative resistance change on the first test trial was significant beyond the .002 level, and a similar test on average relative resistance change over all seven test trials reached the .02 level of significance.[1] Moreover, 12/14 *S*s in Group 1,500 showed a larger absolute GSR on the first test trial than on the pretest trials ($p < .01$, sign test) whereas in Group 40 only 8/15 showed such an increase.

The apparent downward trend of the GSR curves did not approach conventional levels of statistical significance on the Wilcoxon *T* test when each group was tested separately or when both groups were combined. Similarly, the interaction between groups and trials did not approach significance.

Results essentially the same as those presented in Fig. 1 were also found for *mean* relative resistance change, the only difference being that the means were somewhat less stable from test trial to test trial. Similarly, using a measure previously employed by Jones (1961), Champion and Jones (1961), and Champion (1962) which involves plotting the percentage of *S*s in each group which give relative GSR responses on each trial greater than the overall median of all *S*s on all test trials yields results paralleling closely those presented in Fig. 1.

DISCUSSION

Taken as a whole, these data cast some doubt on the generality of Jones' two-process interpretation of classical aversive conditioning. The failure to find any evidence for backward conditioning of either of the short-latency responses at either the 40- or the 1,500-msec. USC-CS interval is consistent with Champion's (1962) recent failure to find backward eyelid conditioning at either 1,200- or 2,800-msec. UCS-CS intervals. When these results are taken in combination with the classic results of Spooner and Kellogg (1947), there is little option but to conclude that backward conditioning of either the eyelid or the finger withdrawal is not possible. This, then, implies that at least that part of Jones' theory relevant to the relationship between backward conditioning and UCR latency is in need of modification.

---

[1] For a given *S*, relative resistance change on each test trial was computed as the ratio of resistance change in ohms on that trial to the average resistance change for that *S* on the three CS-alone trials given immediately prior to the beginning of conditioning. The median of this ratio across *S*s is what is plotted in Fig. 1.

There is left, however, the possibility that this theory does describe what we call conditioning of the GSR. The present study shows increased GSR following just two backward-conditioning trials with a 1,500-msec. interval, but no such increase with a 40-msec. interval. Champion (1962) has recently reported similar results, finding a much larger increase in GSR following three backward-conditioning trials with a 1,200-msec. interval than with a 2,800-msec. backward interval or with a 400-msec. forward interval, and Champion and Jones (1961) have shown backward conditioning of the GSR with a 750-msec. interval in a design controlling for pseudoconditioning. Therefore, it would seem that not only does backward conditioning occur with the GSR but also that the amount of such conditioning is related to the UCR latency in the manner predicted by Jones' theory. The exact implications of this difference between GSR conditioning and the conditioning of other responses are not entirely clear on the basis of research to date, but the possibility that what we call conditioning of the GSR involves a different mechanism from that involved in the conditioning of these other responses cannot be ignored.

REFERENCES

CHAMPION, R. A. Stimulus-response contiguity in classical aversive conditioning. *J. exp. Psychol.*, 1962, **64**, 34–39.

CHAMPION, R. A., & JONES, J. E. Forward, backward, and pseudoconditioning of the GSR. *J. exp. Psychol.*, 1961, **62**, 58–61.

JONES, J. E. The CS-UCS interval in conditioning short- and long-latency responses. *J. exp. Psychol.*, 1961, **62**, 612–617.

JONES, J. E. Contiguity and reinforcement in relation to CS-UCS intervals in classical aversive conditioning. *Psychol. Rev.*, 1962, **69**, 176–186.

SPENCE, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

SPOONER, A., & KELLOGG, W. N. The backward conditioning curve. *Amer. J. Psychol.*, 1947, **60**, 321–334.

(Received June 14, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 392–397

FORMS OF RELEVANT STIMULUS REDUNDANCY IN CONCEPT IDENTIFICATION[1]

ROBERT C. HAYGOOD AND LYLE E. BOURNE, JR.

University of Utah

This study explored the effects of 2 forms of relevant stimulus redundancy within a concept-identification problem. Form A redundancy exists when the levels of 2 stimulus dimensions are perfectly correlated. Form B redundancy exists when the levels of 1 dimension are contingent upon a combination of levels within 2 or more other dimensions, but there is no correlation between pairs of dimensions. The results showed that redundancy of both forms within relevant dimensions facilitated performance relative to problems having the same stimulus uncertainty. Increasing amounts of irrelevant information degraded performance and did not interact with form of redundancy.

The population of stimulus patterns employed in the typical concept-identification study is generated by a fixed set of dimensions, each with two or more levels. If fewer patterns than the total number possible are used in a problem, the population of patterns is redundant. Garner (1962) made explicit the distinction between form and amount of redundancy, defining amount as the number of patterns actually used relative to the total generated by the set of dimensions and form as the particular patterns included.

Bourne and Haygood (1959) have reported evidence on the effects of amount of redundancy of two different forms on performance in concept-identification tasks. When the redundancy is limited to those dimensions which are relevant to problem solution, performance improves as redundancy increases, the effect being more marked as the number of nonredundant irrelevant dimensions increases. The general result holds even in problems with no irrelevant dimensions (Bourne & Haygood, 1960) and despite the fact that relevant dimensions were *added* to the original set of dimensions to create the redundant stimulus population. In another experiment, the amount of irrelevant stimulus redundancy was varied with a resulting decrement in performance on both two-response (one relevant dimension) and four-response (two nonredundant relevant dimensions) problems. The effect of adding redundant irrelevant dimensions was, however, not as great as that attributable to comparable amounts of nonredundant irrelevant dimensions.

In these experiments, any two redundant dimensions were completely correlated, i.e., in any pattern a given level on one dimension was always combined with a particular level on the other. The difference in form of redundancy between these two experiments is based on the relevancy of the dimensions involved. As Garner (1962) points out, there are many other forms of redundancy which may produce differential effects on performance, irrespective of any concomitant variation in the amount of redundancy. There are, however, no experiments which have been purpose-

---

[1] This research was supported in part by a research grant (M-1759) from the National Institute of Mental Health, United States Public Health Service.

fully and adequately designed to study these forms.

Results of an experiment reported by Bricker (1955) suggest that the form of redundancy within relevant dimensions may significantly determine the efficiency of pattern identification. The basic set of eight nonredundant patterns used in Bricker's procedure was generated from three binary dimensions. One and 2 bits of stimulus redundancy were introduced into this set by adding a fourth and a fifth variable dimension while holding the total number of patterns to eight. Results showed that redundancy impeded performance, measured both in rate of learning to assign proper nonsense-syllable labels to the patterns and in speed of responding. Inspection of the patterns selected from the total set that could be generated from five binary dimensions indicates that the two redundant dimensions differed in form. The first redundant dimension, added in all cases to produce 1 bit of redundancy, was uncorrelated with any single dimension but was perfectly predictable from the combinations of levels of three other dimensions. The second redundant dimension, added to produce 2 bits of redundancy, was perfectly correlated with one dimension of the original set of three. Note that such a redundant dimension has precisely the same form used in all conditions of the Bourne and Haygood experiments, wherein relevant redundancy was found to improve reliably the efficiency of concept identification.

In a sense, the variables of form and amount of redundancy were confounded in the design of Bricker's study. Despite this, there is suggestive evidence in the results that form of redundancy is an effective variable irrespective of the amount. Although both redundancy groups produced worse performance than the control, the 2-bit redundancy condition was superior on all measures to the 1-bit condition. In other words, adding the fifth redundant dimension improved performance somewhat, an effect which might be expected on the basis of the experiments of Bourne and Haygood.

A rigorous comparison of the forms of redundancy appearing in the Bricker experiment is warranted. Bricker's procedure does not satisfy two requirements necessary for such a comparison. First, as noted, its design does not provide for independent manipulation of form and amount of redundancy. Second, the proper control conditions for the comparison are missing. Just as in the Bourne and Haygood study, redundancy was introduced by adding dimensions to the initial set while holding the number of different patterns used constant. This has the effect of increasing, with redundancy, the uncertainty in the system of stimulus dimensions which may be an independent determiner of performance (Garner, 1962). For complete information on the effects of stimulus redundancy, a control condition with stimulus uncertainty equal to that for all redundancy conditions is needed.

The purpose of this experiment is to compare the two forms of relevant stimulus redundancy, used by Bricker, within four-response concept-identification tasks. Conventionally, such problems have a unique solution in two binary dimensions, produced by the four conjunctive combinations of levels within these relevant dimensions. To produce the two redundancy conditions, a third dimension was added either by correlating its levels with the levels of one of the two initially relevant dimensions, as with

Bricker's fifth dimension (Form A), or by correlating its levels with a combination of both initially relevant dimensions, as with Bricker's fourth dimension (Form B). Table 1 shows schematically the arrangement of three relevant dimensions under Forms A and B redundancy. For comparison, two control conditions were employed. For one, the third dimension which had been used as redundant was allowed to vary independently, thereby producing stimulus uncertainty equal to that for redundancy conditions. Compared to redundancy conditions, there were twice as many different patterns in the population on which the concept problem is defined in this control. A second control, comparable to that used by both Bricker and Bourne and Haygood, eliminated the third dimension, so that the uncertainty in the system was reduced by 1 bit, but the total number of patterns in the population was the same as in redundancy conditions.

In the face of conflicting results on the effect of redundancy and of somewhat different task requirements in earlier experiments, it is difficult to make any firm predictions on the outcome of this study. For example, Bricker found both forms of redundancy to impede performance relative to a control group, for which the stimulus patterns had lesser uncertainty. With uncertainty constant, redundancy might have facilitated performance in his task. Moreover, Bricker exposed patterns to S for only .7 sec. Garner (1962) argues that short exposures may reduce the influence of relevant stimulus redundancy. Available data do suggest, however, that performance under Form A redundancy will be significantly superior to that under Form B and that, with unlimited time to inspect and respond to a pattern, which is the ordinary condition in concept-identification studies, both forms of redundancy will be superior to a control with equal stimulus uncertainty.

TABLE 1

ARRANGEMENT OF LEVELS WITHIN TWO INDEPENDENT AND ONE REDUNDANT RELEVANT BINARY DIMENSIONS IN FOUR-RESPONSE PROBLEMS

Response	Independent Relevant Dimensions[a]		Redundant Dimension	
	I	II	Form A	Form B
A	1	1	1	1
B	1	0	1	0
C	0	1	0	0
D	0	0	0	1

Note.—Binary numbers designate the different levels within any dimension.

[a] The typical four-response problem has a single unique solution in these two dimensions.

To increase the range of validity of any conclusions that may be drawn from this study, redundancy and control conditions were established under three amounts of irrelevant information. From earlier studies of concept identification (Bourne & Haygood, 1959) it may be expected that any effect of redundancy will be magnified in problems with greater amounts of irrelevant information.

METHOD

Task, procedure, and apparatus.—The task and procedure were essentially the same as those used in earlier concept-identification studies (e.g., Bourne & Haygood, 1959). Each S was presented at the outset with detailed instructions describing (*a*) the problem, including various solution possibilities provided by the presence of redundancy, (*b*) the response requirements, (*c*) the meaning of the feedback signal lamps, (*d*) the stimulus dimensions and their two levels, and (*e*) the criterion of problem solution. The S was required to respond to each of the serially presented geometric stimulus patterns by pressing one of four buttons. Immediately after each response, the pattern was removed and a feedback lamp was turned on over the correct button. The *S*'s task was to learn that these buttons corresponded to the conjunctive combinations of levels of the dimensions relevant to problem solution. One or more irrelevant dimensions, the levels of which were uncorrelated with correct responses, appeared in all problems. As shown in Table 1, when either form of redundancy was present in a problem more than one solution or labeling scheme for the set of four buttons was correct. In all cases, however, only one response was correct for any given pattern. The criterion of problem solution was 16 consecutive correct responses. All conditions were self-paced in that *S* was allowed as much time as needed to respond to any pattern.

The presentation of the patterns, the control of feedback signals (1 sec. duration) and of overall interstimulus interval (5 sec.), and the recording of responses were accomplished automatically with the concept-identification apparatus described by Bourne and Haygood (1959).

Subjects and design.—The *S*s were 288 students from introductory psychology classes who received class credit for participation. They were assigned randomly and in equal number to the 24 independent conditions of the experiment. In addition, 11 *S*s were eliminated for a variety of reasons, including failure to solve and temporary equipment malfunction. These *S*s were distributed without bias across groups.

The experimental design was a 4 × 3 × 2 factorial, with (*a*) four types of relevant information: Form A and Form B redundancy, nonredundant relevant information with the third dimension (redundant in Forms A and B) eliminated (Control I), and nonredundant information with the third dimension present but irrelevant (Control II); (*b*) three levels of task complexity: one, two, and three irrelevant dimensions; and (*c*) two different problems. Note that in Control II, the numbers of irrelevant dimensions were actually two, three, and four. Two groups of Control II (two and three irrelevant dimensions) overlapped Control I except for the particular stimulus dimensions chosen to be irrelevant. These conditions were included because of known differences in the obviousness of various dimensions (Archer, 1962; Brown & Archer, 1956).

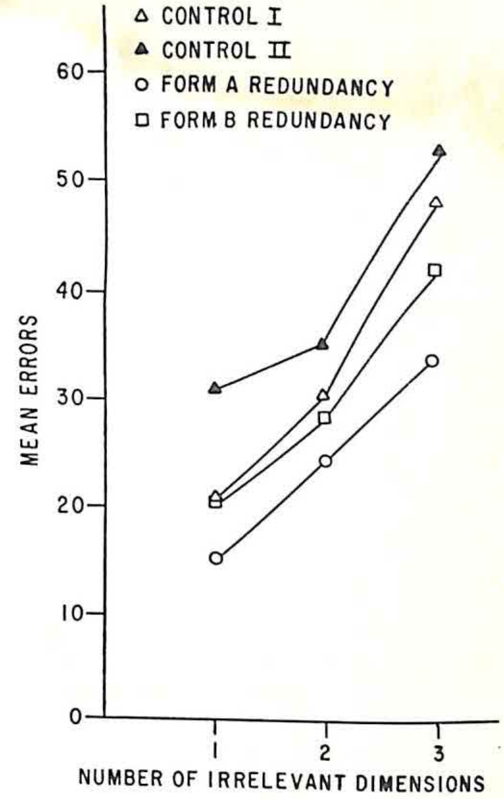


FIG. 1. Mean errors as a function of number of irrelevant dimensions and relevant-information condition. (Each plotted point represents 24 *S*s.)

RESULTS AND DISCUSSION

Mean number of errors to criterion as a function of type of relevant information and number of irrelevant dimensions is shown in Fig. 1. An analysis of variance performed on these data showed reliable overall differences among conditions of relevant information, $F(3, 264) = 5.46$, $p < .005$. Moreover, according to Duncan's test, performance under Form A redundancy was significantly superior to that under either control condition ($p < .05$). Performance under Form B redundancy, although numerically superior to Control I, was not significantly different. As expected, performance in Control II was worst; Form B redundancy was clearly superior to this condition ($p < .05$). Further analyses comparing the two forms of redundancy showed Form A superior to Form B, $F(1, 132) = 4.14, p < .05$. A similar analysis of trials to solution showed exactly the same outcome.

The results thus indicate that relevant stimulus redundancy facilitates performance in four-response concept-identification problems. The outcome is consistent with the results of Bourne and Haygood (1959) who found Form A redundancy to facilitate performance in similar two-response tasks. Moreover, the reliable difference between two forms of redundancy confirms Garner's (1962) expectation about the importance of this as a variable, independent of amount of redundancy.

These results appear to be discrepant with Bricker's (1955) observation of a deleterious effect of relevant stimulus redundancy on pattern identification. While the difference in outcomes might be explicable on the basis of numerous differences in task requirements between experiments, in the interest of parsimony it is reasonable to search for an explanation which unifies these data. The following four considerations are relevant to such an interpretation. (*a*) Bricker's experiment lacked one important control. His two redundancy conditions might have been superior to nonredundant conditions with comparable amounts of stimulus uncertainty, i.e., with the fourth and fifth dimensions varying independently as irrelevant dimensions. Data of the present experiment show the facilitative effect of redundancy to be reduced when comparison is made against a control, analogous to Bricker's, with the same total number of patterns but less uncertainty. (*b*) Even against this control, however, redundancy improved performance in the present experiment, whereas it seemed to impede performance in Bricker's study. For this reason, it is important to take cognizance of the nature of the stimulus displays used. Adding a (redundant) dimension to Bricker's display entailed the inclusion of a spatially separate binary element and thus increased the complexity of the display. Perceptually this means that *S* had to scan a broader field as the number of dimensions increased. In the present tasks, stimulus patterns were relatively compact (Shepard, Hovland, & Jenkins, 1961) so that the addition of a dimension did not increase the complexity of any single pattern, although the amount of variation across patterns was greater. (*c*) Under conditions of unlimited exposure, such a difference in stimulus displays may not be important; there is little evidence on the relative difficulty of perceiving information in discrete and compact patterns. However, Bricker, probably because his study was aimed at assessing reaction time as well as identification learning, exposed each pattern for only .7 sec. Under such circumstances, as Garner (1962) argues, the redundancy may not have existed, perceptually, for *S*. The *S* may not have seen all elements of any given pattern. If this is the case, inferior performance with redundancy would be expected because *S* can only use certain combinations of dimensions to respond correctly to redundant stimulus patterns. (*d*) Finally, in Bricker's experiment, the difference between forms of redundancy was confounded with amount of redundancy. In the present experiment, these two forms of relevant redundancy were isolated, with the result that Form A facilitated performance more than Form B. Taking into account the above-mentioned considerations, Bricker's data are quite consistent with this finding. He observed the fifth (second redundant) dimension to improve performance over the 1-bit redundancy condition. It may be argued that the expected facilitation by Form A redundancy was sufficient to overcome interference resulting from short stimulus exposure and complexity and thus to produce better performance than that with Form B alone. In this way, both experiments, plus the earlier studies of Bourne and Haygood, argue that Form A is a powerful source of relevant information to a learner.

Why Form A redundancy is superior to Form B remains to be answered.

There is at least one reason to expect problems with Form B redundancy to be easier. From Fig. 1, it can be seen that Form B provides three mutually independent ways in which a problem can be solved whereas Form A redundancy gives only two. The most reasonable interpretation for the superiority of Form A lies in Garner's (1962) hypothesis that not all forms of redundancy or structure within a system of stimulus dimensions are equally perceivable. Form A redundancy is based on a perfect correlation between the levels of the redundant dimensions and a single relevant dimension. In Form B, no single relevant dimension correlates with the redundant dimension, but rather the redundancy exists in the interaction of the levels of the two relevant dimensions and the redundant dimension. These data are in agreement with Garner's contention that Form B redundancy is much less easily perceived and utilized than Form A. The facilitative effect of Form B redundancy on concept identification is clearly reliable, however, at least when compared to a control condition with the same amount of stimulus uncertainty.

As is typical in concept identification (Bourne & Haygood, 1959), the effect of number of irrelevant dimensions on errors to solution proved to be significant, $F(2, 264) = 21.55$, $p < .001$, and essentially linear, $F_{\text{lin.}}(1, 264) = 41.35$, $p < .001$. No difference was found between problems, nor, despite the apparent lack of parallelism among functions shown in Fig. 1, were any of the interactions significant. An interaction between amount of irrelevant information and condition of relevant information was expected on the basis of earlier studies of stimulus redundancy in concept identification. However, in the earlier experiments, amount rather than form of redundancy was the manipulated, interacting variable. In the present study, only 1 bit of redundancy was permitted; in fact, only 1 bit of Form B redundancy is possible within the four-response problem although the amount of Form A is unlimited. It may be that differing amounts of Form A redundancy in such tasks would interact with number of irrelevant dimensions and that, in problems with other response structures, differing amounts of Form B redundancy would interact with irrelevant information. Such a demonstration was not the purpose of the present investigation, however. There is evidence here that form of redundancy, in limited amounts, does not interact with irrelevant information and that the conclusions concerning the relative efficacy of forms of redundancy need not be qualified on this variable.

REFERENCES

ARCHER, E. J. Concept identification as a function of obviousness of relevant and irrelevant information. *J. exp. Psychol.*, 1962, **63**, 616–620.

BOURNE, L. E., JR., & HAYGOOD, R. C. The role of stimulus redundancy in concept identification. *J. exp. Psychol.*, 1959, **58**, 232–238.

BOURNE, L. E., JR., & HAYGOOD, R. C. Supplementary report: Effect of redundant relevant information upon the identification of concepts. *J. exp. Psychol.*, 1960, **61**, 259–260.

BRICKER, P. D. The identification of redundant stimulus patterns. *J. exp. Psychol.*, 1955, **49**, 73–81.

BROWN, F. G., & ARCHER, E. J. Concept identification as a function of task complexity and distribution of practice. *J. exp. Psychol.*, 1956, **52**, 316–321.

GARNER, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

SHEPARD, R. N., HOVLAND, C. I., & JENKINS, H. M. Learning and memorization of classifications. *Psychol. Monogr.*, 1961, **75**(13, Whole No. 517).

(Received June 19, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 400–402

EFFECT OF INSTRUCTIONS UPON EYELID CONDITIONING[1]

MARGARET F. NICHOLLS AND GREGORY A. KIMBLE

Duke University

2 groups of 16 Ss each received 40 eyelid conditioning trials with facilitative and inhibitory instructions. Confirming previous findings, it was shown that the latter instructions inhibited the progress of conditioning. Supplementing previous results, this effect was shown to be associated with habituation of the UCR.

Studies by Miller (1939) and Norris and Grant (1948) have demonstrated that instructions have an important effect upon eyelid conditioning: So-called "inhibitory" instructions, by comparison with "facilitative" or "neutral" instructions, result in a lowered level of conditioning. The purpose of this study was to test the hypothesis that this effect is the result of habituation of the UCR which, in a very different context, Kimble and Pennypacker (1963) have shown to be associated with poor conditioning.

Method.—Two groups of 16 undergraduate students each were conditioned with facilitative and inhibitory instructions. There were 8 males and 8 females in each group. The *S*s were assigned to conditions alternatingly in order of appearance until it became necessary to depart from this procedure for the purpose of equating groups with respect to sex.

The eyelid-conditioning apparatus was a modification of that previously used in this laboratory (Dufort & Kimble, 1958). The *S* sat in a well-lighted, sound-deadened chamber. Ink-written records of the eyeblink were obtained by a system consisting of a microtorque potentiometer connected to a false eyelash taped to *S*'s right eyelid, a bridging circuit supplied by the Grass Instrument Company, a Grass driver amplifier, and a Grass Model 5 polygraph. The UCS was a .10-sec. air puff delivered through a $\frac{3}{64}$-in. aperture to the right cornea from a distance of about .5 in. at a pressure capable of supporting a 180-mm. column of mercury. The intertrial interval varied irregularly from 10 to 20 sec. averaging 15 sec. The CS was the onset of a circular red light of $1\frac{1}{4}$ in. diameter, produced by a Grason-Stadler multiple-stimulus projector. The CS-UCS interval was .5 sec.

The two sets of instructions were identical except that *S*s given facilitative instructions were told to: "Relax and let your eye reactions take care of themselves. If you feel your eye close or about to close, do nothing to stop it." The *S*s given inhibitory instructions were told to: "Concentrate on not blinking until you feel the puff of air. That is, try not to blink after the light comes on until you feel the air puff."

After receiving these instructions, *S*s in both groups received 5 preliminary trials (2 trials with the CS alone followed by 3 trials with the UCS alone) and 40 paired CS-UCS conditioning trials.

Results.—All responses with latencies from .25 to .50 sec. and a deflection of 1 mm. or more from a stable baseline were counted as CRs. Conditioning curves for the groups with inhibitory and facilitative instructions appear in Fig. 1 where percentages of CRs are plotted against trials in blocks of 5. The difference in level of conditioning between the two groups was statistically significant for the entire 40 trials and for the last 20 trials. Respective t values ($df = 30$) were 4.20 and 3.41 ($p = .005$ in each case).

Before testing our hypothesis that the inferior conditioning of the group receiving inhibitory instructions was associated with differential *habituation* of the UCR, it is important to show that the differences in instructions produced no difference in the *initial* strength of the UCR. Accordingly, median amplitudes of the UCR on the preliminary trials with the UCS alone were determined for each *S*. The means of these values, 34 mm. and 33 mm. for the facilitative and inhibitory groups, respectively, were not reliably different ($t = 0.32$). Thus there is no evidence that the different instructions had an immediate effect upon the amplitude of the UCR.

To test the hypothesis that the effect of instructions upon conditioning might have been produced by differential habituation, two analyses were performed. The first involved a comparison of the amplitudes of the UCR on the last conditioning trial on which no CR occurred. The mean of these magni-

---

[1] This experiment received financial support from National Science Foundation Grant NSF-G 7079.

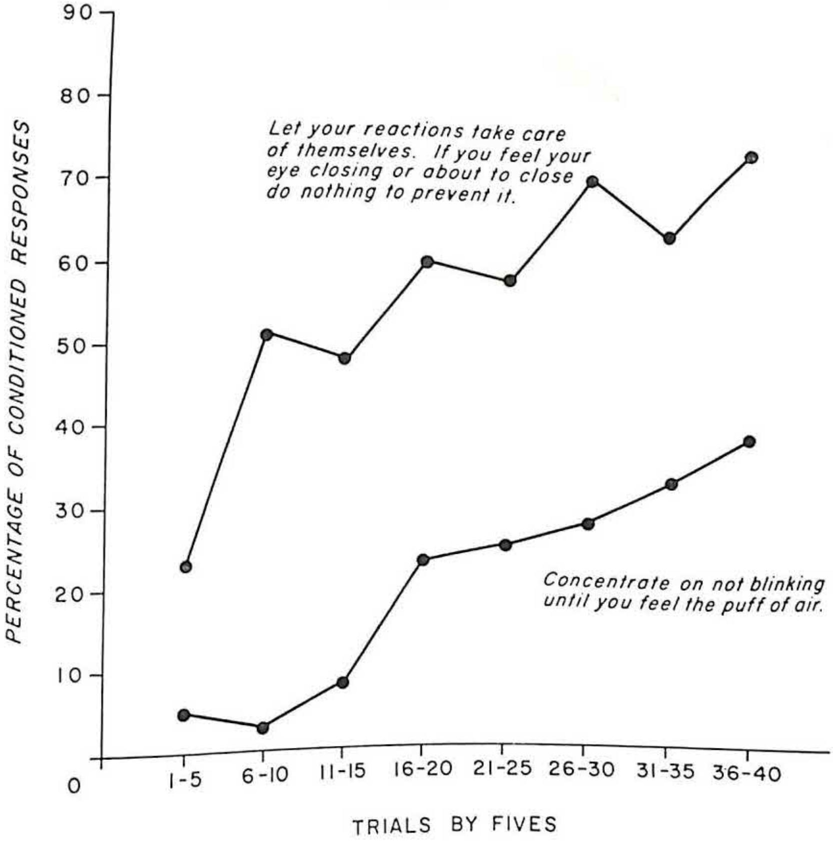


FIG. 1. Mean percentage of CRs for *Ss* conditioned with facilitative and inhibitory instructions.

tudes was 26 mm. for the group receiving inhibitory instructions and 31 mm. for the group receiving facilitative instructions. The difference approaches significance, t (30) = 1.8, p = .10. The second analysis supplies clearer evidence in support of the hypothesis. This was an analysis of the amount of decrease in UCR amplitude during conditioning as reflected by the difference in amplitude of the UCR before conditioning and on the last conditioning trial without a CR. The respective decreases were 2.69 mm. for the facilitative-instructions group and 8.31 mm. for the inhibitory-instructions group, t (30) = 2.06, p = .025. The increase in significance between the previous analysis and this one presumably reflects the precision gained in the second analysis by taking initial response strength into account.

The results of this experiment suggest the following conclusions: (*a*) Confirming the results of previous studies, facilitative and inhibitory instructions result in different levels of conditioning. (*b*) This difference is associated with the greater degree of habituation of the UCR in *Ss* receiving inhibitory instructions.

As a further test of the second conclusion, conditioning was compared within each group for *Ss* who showed large and small amounts of habituation. Each major group of *Ss* was divided into two subgroups. The two groups were those *Ss* whose decreases in UCR amplitude were above and below the median difference between the pre-experimental UCR amplitude and the UCR amplitude on the last trial without a CR. For the group conditioned with inhibitory instructions, the mean numbers of CRs for the subgroups showing large and small amounts of habituation were 4.6 and 11.5, respectively. Evaluated by a one-tailed t test, the difference between these values was significant, t (14) = 1.86, p = .05. For the subgroups with facilitative instructions, the trend was in the same direction. Comparable mean numbers of CRs were 19.0 and 25.6. The difference, however, was not significant (t = 1.2).

A point of some importance is that the results of this experiment almost exactly parallel those of Kimble and Pennypacker in their comparison of eyelid conditioning in young and aged *S*s. The suggestion is that the degree of habituation of the UCR is a process of general significance in determining the progress of this form of learning. It seems likely, however, that, whereas differences in habituation resulted from physiological differences in the Kimble and Pennypacker study, in this study they reflect differences in set or attitude.

REFERENCES

DUFORT, R. H., & KIMBLE, G. A. Ready signals and the effect of UCS presentations in eyelid conditioning. *J. exp. Psychol.*, 1958, **56**, 1–7.

KIMBLE, G. A., & PENNYPACKER, H. S. Eyelid conditioning in young and aged subjects. *J. genet. Psychol.*, 1963, **103**, 283–289.

MILLER, J. M. The effect of inhibitory and facilitatory attitudes on eyelid conditioning. *Psychol. Bull.*, 1939, **36**, 577–578. (Abstract)

NORRIS, E. B., & GRANT, D. A. Eyelid conditioning as affected by verbally induced inhibitory set and counter reinforcement. *Amer. J. Psychol.*, 1948, **61**, 37–49.

(Received June 1, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 4, 402–404

A REPLICATION OF SELECTIVE ATTENTION AND CODING IN VISUAL PERCEPTION[1]

RALPH NORMAN HABER

Yale University

The replication supported and extended the findings of an experiment by Harris and Haber (1963). The form of Ss' verbal encodings of briefly presented stimuli was found to account for the effects of attention instructions on accuracy of their reports. Attention instructions had a smaller effect on Ss who encoded the stimulus in the order dictated by English syntax. The Ss who encoded first the stimulus attribute they were told to attend to, reported this attribute more accurately.

Harris and Haber (1963) reported that the effects of selective attention on a perceptual task could be adequately accounted for by differences in the way *S*s verbally encoded the stimulus before reporting it. Further, checking on a contention of Lawrence and LaBerge (1956) that the effects of attention instructions were caused by selective ordering of reports, Harris and Haber showed that sets induced by attention instructions produced differential accuracy of report even when the order of report was fully controlled by *E*. All of their results were interpreted in terms of memory processes, particularly the rapid decay of some short-term immediate memory, and in terms of different coding strategies used by *S*s—codes that probably differed in the speed with which the stimulus could be encoded, and in the resistance to interference during the subsequent report of the stimulus.

Method.—The following experiment represents a complete replication of Harris and Haber's study. The same population, stimuli, presentation procedures, training procedures, and response instruments and analysis procedures were used.

Each of the 80 stimuli consisted of a pair of Wisconsin Sorting Cards. They were flashed for 1/10 sec., after which *S* had from 15 to 20 sec. to record what he had seen. The cards differed in the number of figures portrayed, their shapes, and their colors. Each of these three dimensions had four values. Before the experimental session began, each *S* was trained to use one of the two basic coding strategies to remember the stimuli while reporting them: *Objects code*, which describes the stimulus as two sets of objects (e.g., "two red circles, four blue stars"), or *Dimensions code*, which describes the three dimensions of the stimulus (e.g., "red, blue; two, four; star, circle").

Before each stimulus was flashed, *S* read instructions about what to attend to. *Emphasis instructions* told *S* to pay special attention to one of the three dimensions of the stimulus (emphasized dimension) though he still had to report the other two dimensions (unemphasized dimensions). *Equal instructions* told *S* to pay equal attention to all three dimensions.

---

[1] This research was supported in part by a grant (MH-03244) from the United States Public Health Service to the author. Grateful acknowledgment is given to Donald Richard and Michael Ross, who helped gather the data, to Martha Breen, who helped with the analysis, and to Charles S. Harris, for his assistance in reading the manuscript.

Order of report was not known to *S* until after each stimulus was presented. On half of the trials, the order in which the three dimensions were to be reported was specified by *E* after the flash (forced order). On the remainder of the trials, *S* could report them in any order he pleased, though he had to indicate the order he used (free order).

Results.—Of the 49 *S*s tested, 41 followed a consistent verbal-encoding strategy throughout the experiment, as classified by the post-experiment questionnaire. The same four codes were found here as in the original experiment: an Objects code, a Dimensions code, a Modified-Objects code, where the attended-to attribute is emphasized in an otherwise Objects code; and an Incomplete-Dimensions code, where not all attributes are encoded. Seven *S*s used a strategy opposite to that for which they were trained, and they, along with the 8 *S*s who did not use a single strategy consistently were dropped from further analyses. The remaining 34 *S*s were those who developed a consistent code similar to the one for which they were trained.

A large main effect of coding strategy was found in the replication as in the original experiment ($p < .01$). Unlike the original study, all four groups are significantly different from each other. (Each of the six possible *t* tests were significant at at least the .01 level.) Therefore, this experiment succeeded in producing a difference in the two basic codes, when used in their unmodified form, as well as differences in their modified versions.

A main effect of set instructions was found, in the same direction as in the previous study ($p < .001$); emphasized instructions produced greater accuracy than unemphasized instructions, equal emphasis instructions between these. The higher error rate for unemphasized dimensions occurs on the first report under forced order of report ($p < .01$), so it cannot be due to some decay of memory over time, and it cannot be due to *S*'s chosing to report first the dimensions he was most certain about. Therefore, one of the major findings of the first experiment is supported. Also, as in the previous study, coding strategy interacts with emphasis instructions ($p < .05$), so that the two Objects groups show smaller effects of set than do the two Dimensions groups. However, in the earlier study, Objects coders reported unemphasized dimentions about as accurately as emphasized (with both free and forced order of report). In this replication, on the other hand, the Objects coders reported emphasized dimensions more accurately than unemphasized with forced order of report ($p < .05$) (though not with free order of report). Any hypothesis about what produces a set effect, whether it be due to perception, coding, interference, or decoding processes, would predict that the set effect would be *greater* under free order of report. Thus, especially in the light of the earlier findings, the Objects coders' set effects under free order of report are suspect.

Table 2 presents the order of report effects. A main effect is again shown ($p < .001$), indicating that items reported early are more likely to be correct than items reported later. This effect is greater on free than on forced order of report ($p < .01$), presumable because the Dimensions coders deliberately report the unemphasized dimensions last when free to do so. (With free order of report, both Dimensions groups reported the emphasized

TABLE 1

MEAN PERCENTAGE OF ERROR

	Free Order of Report			Forced Order of Report			*M*
	Emphasized Dimensions	Unemphasized Dimensions	Equal Dimensions	Emphasized Dimensions	Unemphasized Dimensions	Equal Dimensions	
Objects coders ($N = 9$)	12.1	12.9	8.1	6.6	12.0	9.5	10.2
Dimensions coders ($N = 12$)	8.9	19.5	14.5	9.5	20.2	14.5	14.4
Modified objects ($N = 9$)	12.1	16.6	12.3	8.4	15.6	11.9	12.8
Incomplete-Dimensions coders ($N = 4$)	15.6	33.1	24.0	11.5	27.6	21.9	22.3
Total ($N = 34$)	11.4	18.6	13.3	8.7	17.7	13.2	13.8

Note.—Positions of stimulus (left and right) are not shown to conserve space.

Bi-monthly

$20.00 a year
($20.50 foreign)
$4.00 a copy

JOURNAL OF COMPARATIVE & PHYSIOLOGICAL PSYCHOLOGY

Editor
W. K. Estes
Stanford University
Associate Editor
Eliot Stellar
University of Pennsylvania
Consulting Editors
Gordon H. Bower
Stanford University
Joseph V. Brady
Walter Reed Army Institute of Research
Byron A. Campbell
Princeton University
Vincent G. Dethier
University of Pennsylvania
Frank A. Geldard
Princeton University
John F. Hall
Pennsylvania State University
Margaret Kuenne Harlow
Wisconsin Regional Primate Research Center
John I. Lacey
Fels Research Institute
Douglas H. Lawrence
Stanford University
Daniel S. Lehrman
Rutgers University
Frank A. Logan
Yale University
Mortimer Mishkin
National Institute of Mental Health
Conrad G. Mueller
Columbia University
Carl Pfaffmann
Brown University
W. N. Schoenfeld
Columbia University
Larry Stein
Wyeth Laboratories
Philip Teitelbaum
University of Pennsylvania
Hans-Lukas Teuber
Massachusetts Institute of Technology
J. M. Warren
Pennsylvania State University
David Zeaman
University of Connecticut

CONTENTS

Weiskrantz, Lawrence, and Cowey, Alan. Striate Cortex Lesions and Visual Acuity of the Rhesus Monkey........ 225

Gross, Charles G. Locomotor Activity Following Lateral Frontal Lesions in Rhesus Monkeys.................... 232

Wilson, William A., Jr., Oscar, Marlene, and Gleitman, Henry. The Effect of Frontal Lesions in Monkeys upon Widely Spaced Delayed-Response Trials................ 237

Warren, J. M., Grant, R., Hara, K., and Leary, R. W. Impaired Learning by Monkeys with Unilateral Lesions in Association Cortex.................................... 241

Stamm, J. S., and Knight, Marcia. Learning of Visual Tasks by Monkeys with Epileptogenic Implants in Temporal Cortex.. 254

Thompson, Robert. Thalamic Structures Critical for Retention of an Avoidance Conditioned Response in Rats...... 261

(continued on inside back cover)

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street, N.W.
Washington, D.C. 20036

Journal of
Experimental Psychology

VOL. 67, NO. 5 MAY 1964

DEVELOPMENT AND ATTENUATION OF DELAY-ENGENDERED AVOIDANCE BEHAVIOR

H. WAYNE LUDVIGSON, WILLIAM F. CAUL, JAMES H. KORN
Carnegie Institute of Technology

AND JAMES H. McHOSE
Southern Illinois University

24 male rats received 108 food-rewarded trials under different delay-of-reinforcement conditions in an L-shaped double alley having a delay chamber at the juncture of the 2 alleys. Following an initial increase, speeds approaching the delay exhibited a slow but pronounced decline followed by a recovery of strength late in training. During this latter phase, starting speeds following the delay significantly decreased in magnitude. These results suggested that not all of the effects of delay can be attributed simply to a limiting of the growth of reaction potential, but that at least 1 effect is inhibitory or subtractive. In addition, it appeared that such an inhibitory effect may, under certain conditions, diminish with extensive training.

The effects of a temporal delay of reinforcement have been variously interpreted. While certain earlier views (Hull, 1943; Logan, 1952) postulated lesser excitatory potential (E) as a consequence of delay, more recent conceptions hold that delay has an inhibitory function analogous to extinction, and thus affects behavior through a factor which subtracts from E. A variant of this view considers delay to be one of the variables which produces an emotional reaction of frustration. Consequent decremental effects on responses approaching a delay occur, then, because of competing reactions mediated by anticipatory frustration (Spence, 1960).

One implication of the Amsel (1958) and Spence (1960) formulations of frustration theory is that any effects of delay which are frustration engendered should be minimal or absent early in training. This follows from the assumption that the occurrence of frustration, and consequently anticipatory frustration, r_f, is contingent upon the presence, in some strength, of anticipatory reward, r_g, which itself must be acquired in the initial phase of training.

The evidence concerning delay and frustration theory is inconclusive.

For example, the studies of Holder, Marx, Holder, and Collier (1957) and Wist (1962) provide data consonant with the assumption of an inhibitory or suppressive effect on response strength approaching a delay. However, since these investigations employed an initial series of nondelayed or minimally delayed acquisition trials, the dependence of the inhibition upon prior development of r_g cannot be determined. The data of Logan (1960) tend to suggest a dependence on r_g, but a number of earlier studies (Perin, 1943; Ramond, 1954) require no such assumption, principally because of a very early onset of delay effects.

The present study was concerned with bringing more data to bear upon these theoretical problems. What appeared to be needed was an experimental setup in which frustration effects, if present, would be likely to occur in some strength, and in which delays are imposed from the outset of training. In addition to establishing these conditions, this experiment served as a preliminary investigation into related problems concerned with variable delay and specific behavior occurring during a delay. However, discussion will be limited primarily to the changes in response strength which occurred over trials in all *S*s.

METHOD

Subjects.—Twenty-four experimentally naive male rats of the Sprague-Dawley strain, 90–100 days of age at the beginning of the experiment, served as *S*s.

Apparatus.—The *S*s were run in an L-shaped double runway, consisting of a start box (SB), a straight alley (A_1), a delay box (DB), a second alley (A_2), and a goal box (GB). The first alley was $3\frac{1}{2}$ ft. long, A_2 was 10 ft. long, and the SB, DB, and GB were each 12 in. in length. The maze was 3 in. wide and $5\frac{3}{4}$ in. high throughout. The walls and floor of all sections were white. Fourteen inches above the roof, which consisted of $\frac{1}{2}$-in. hardware cloth, there was a translucent plastic sky which served to diffuse the light from 6-w. 120-v. lamps situated 13 in. above the sky and spaced at a distance of 13 in. A small wooden block with indentations for food pellets served as a food cup in the GB.

Guillotine start and retrace doors separated the five sections of the apparatus. Starting speeds in A_1 and A_2 were calculated on the basis of the time lapsed between the opening of the SB or DB start door and *S*'s tripping a microswitch which was operated by a very slight depression of a trap door located 12 in. from the start door. The time taken by *S* in traversing the next 18 in. of A_1 or 96 in. of A_2 before tripping another microswitch provided the basis for the running-speed measures. All speed measures, expressed in ft/sec, were the means of *S*s' median speeds over a block of trials. A retracing in A_1 was recorded when *S*, within 12 in. of the DB and having tripped the last microswitch in the alley, turned and moved in the direction of the SB.

Procedure.—Sixteen days prior to the first training day, *S*s were placed on a 23-hr. feeding schedule which was maintained throughout the experiment. Water was available at all times except during a trial in the runway. On Days 11–13 of the preliminary period, *S*s were removed from their home cages and held by *E* for 2 min. in order to accustom them to being handled.

To habituate *S*s to the maze, they were allowed 5 min. of exploration in each alley on each of Days 14–16.

Training began on Day 17. On the first 2 days 2 trials were given. Thereafter, *S*s received 3 trials per day until a total of 108 trials had been administered. Within any day the intertrial interval was approximately 20 min. A trial began with the placement of *S* into the SB facing the rear of the box. The first start door was opened after *S* had maintained an orientation toward it for 3 sec. Immediately after *S* entered the DB the guillotine retrace door separating A_1 and the DB was closed. Following a prescribed delay interval, the second start door was opened and *S* was allowed to traverse A_2. When *S* entered the GB the second retrace door was lowered, and *S* was returned to its cage immediately after it had eaten two .045-gm. Noyes pellets.

Three groups received different delay treatments in the DB. Group 10 was always confined for 10 sec. Group V had a variable-delay interval. On half the trials it was delayed for 10 sec. while on the other half it was held 1 sec. according to the following schedule: 1, 10, 1; 10, 1, 10; 1, 1, 10; 10, 10, 1; 1, 10, 10; 10, 1, 1, which was then repeated.

Group 1 always had a 1-sec. delay until Trial 72, after which it was shifted to the variable 1- and 10-sec. schedule of Group V.

RESULTS

Alley 1: Response strength approaching the delay.—Mean starting and running speeds for A_1 are shown in Fig. 1 and 2, respectively. Clear indication of nonmonotonicity of trends appeared in all groups. Following an initial increase in speed there was a protracted period of training in which speeds slowly declined and then tended to recover late in training.

A summary of the analysis of variance of all nine blocks of trials is presented in Table 1. The trials effect, as would be expected, was significant. In order to assess, roughly, the curvature of the overall trend, the linear, quadratic, and cubic components of the trials factor were extracted. The cubic components of both starting- and running-speed measures were significant well beyond the .001 level. For the starting-speed measure the quadratic component was also significant at the .05 level and for the running-speed measure, there was a negative linear trend significant beyond the .001 level. The cubic trends alone accounted for 67 and 50% of the total variance attributable to trials in the starting- and running-speed measures, respectively. Together, the two components which were significant in each case accounted for 85 and 86% of the trials variance.[1] These results indicate that following an initial increase in speed there was a significant

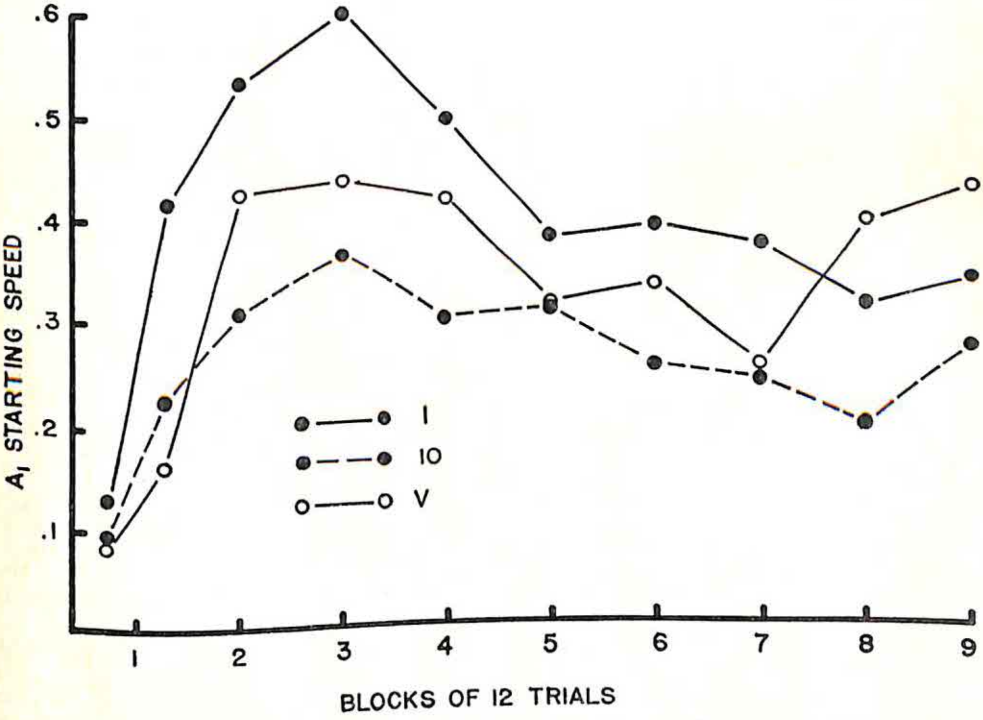


FIG. 1. Mean starting speeds in Alley 1 for Groups 1, 10, and V. (The first block of trials is divided into six-trial points to show initial response strengths.)

---

[1] It should be noted parenthetically that, at least in the case of the starting-speed measure, there was a significant deviation from the least-squares, best-fitting power series having linear, quadratic, and cubic components, which indicates that this type of function does not provide the best possible fit to the data.

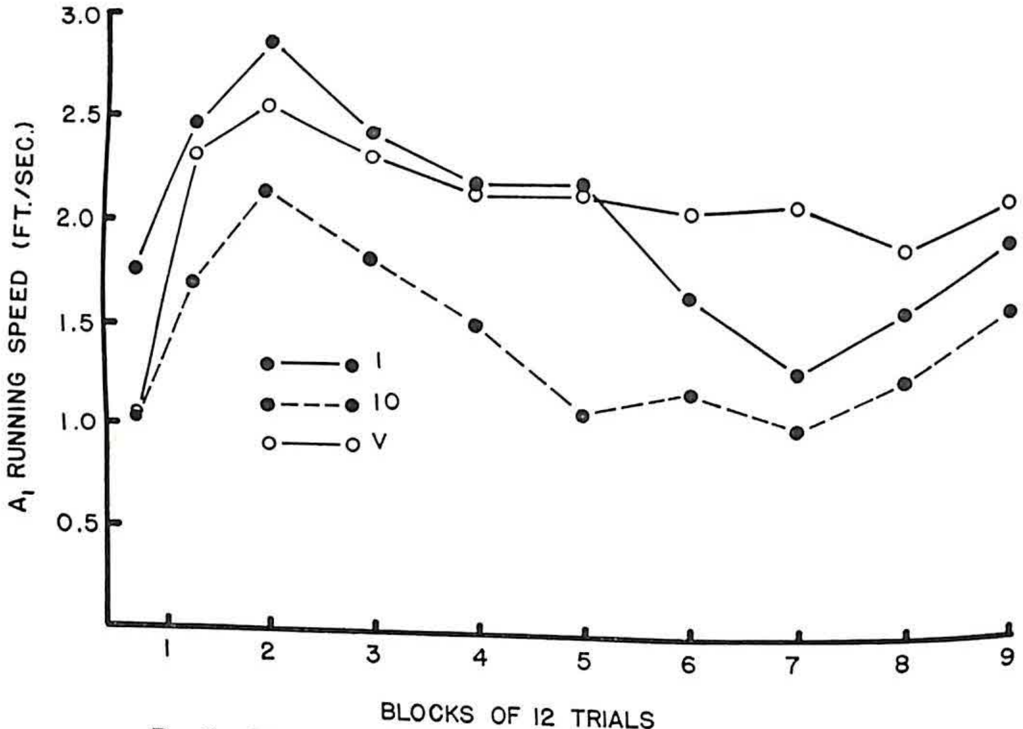


FIG. 2. Mean running speeds in Alley 1 for Groups 1, 10, and V.

decrement and then possibly a second increase in response strength. To evaluate further the changes occurring toward the end of training the last three blocks of trials, 7–9, were analyzed separately. The trials component of the running-speed measure was significant, F (2, 42) = 3.78,

TABLE 1

ANALYSIS OF VARIANCE OF A_1 STARTING AND RUNNING SPEEDS

Source	*df*	Starting Speed		Running Speed	
		MS	*F*	*MS*	*F*
Between *S*s	23				
Groups (G)	2	.3802	1.77	7.603	2.45
*S*s within Gs	21	.2153		3.100	
Within *S*s	192				
Trial blocks (T)	8	.1879	9.50***	3.162	8.66***
Linear	1	.0009	<1	9.045	15.12*
Quadratic	1	.2715	7.85*	.345	<1
Cubic	1	1.0016	38.94***	12.618	38.29***
Residual	5	.0459	4.32*	.658	2.29
T × G	16	.0224	1.13	.573	1.57
T × *S*s within Gs	168	.0198		.365	
Linear	21	.0448		.598	
Quadratic	21	.0346		.556	
Cubic	21	.0257		.330	
Residual	105	.0106		.287	
Total	215				

* $p < .05$.
*** $p < .001$.

TABLE 2

Frequency of Retracing in A_1

Group	Blocks of 12 Trials									Group Totals
	1	2	3	4	5	6	7	8	9	
1	1	1	0	1	3	6	4	1	1	18
10	0	0	4	15	14	16	10	8	2	69
V	1	3	3	14	13	6	3	2	0	45
Block totals	2	4	7	30	30	28	17	11	3	132

$p < .05$. It did not reach conventional levels of significance in the starting speeds, $F(2, 42) = 2.14$, $p > .05$, in which the recovery for Groups 1 and 10 was not so pronounced and appeared not to begin until Block 8.

Table 2 presents the frequencies of retracings which were initiated from a point within 12 in. of the delay chamber. Since these retracings began after *S*s had traversed the segment of A_1 from which running speed was recorded, they constitute data independent from the speed measures. As may be seen, the frequency of retracing correlated quite closely with both the decline and recovery of the speed measures. It should be noted that not all *S*s retraced. In fact, of the total 132 retracings, 107 were contributed by only seven *S*s. Nevertheless, such retracings as did occur were mainly in the middle trials of the study.

As indicated in Table 1, the differences among the three groups did not attain statistical significance nor were there reliable differences among the trends as indicated by the nonsignificant Groups × Trials interaction effects. Although Group 1 remained

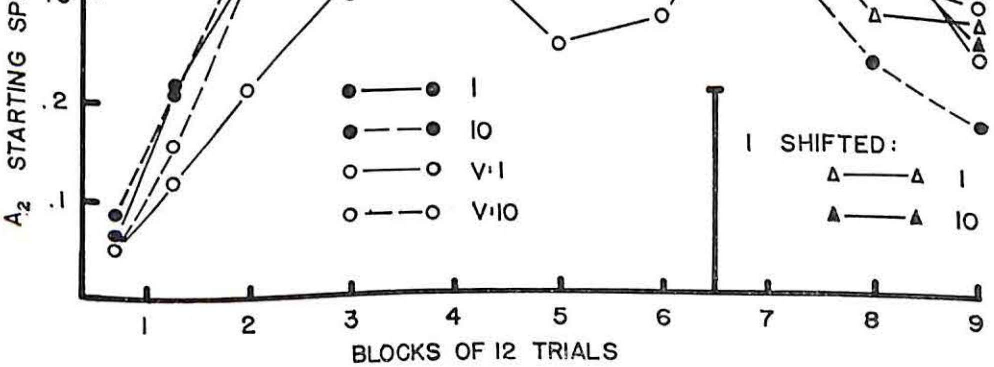


Fig. 3. Mean starting speeds in Alley 2. (Speeds following the 1- and 10-sec. delays within a block are plotted separately for Group V, and for Group 1 after its shift to the variable condition.)

consistently above Group 10 at all points in training, a number of subsequent analyses still failed to reveal significant differences.

Alley 2: Response strength following the delay.—With two exceptions, noted below, the running speeds were very similar to the starting speeds, and therefore, only the latter are presented in Fig. 3.

The most noteworthy feature of the A_2 starting speeds is the overall decrease in response vigor late in training. This is of particular interest since it occurred in a period of training during which A_1 speeds were increasing. Analysis of the data for Trial Blocks 7–9 revealed a significant trial-block effect, $F(2, 42) = 14.06$, $p < .001$. The decrement in the A_2 running speeds was not significant, $F(2, 42) = 1.79$, $p > .10$, indicating that this effect was more reliable in that part of the response chain closest to the delay box.

The considerable intertwining of the curves in Fig. 3 suggests that there were no significant differences among the experimental conditions. Several analyses of variance supported this conclusion. This statement also holds for the running-speed data, with the exception that *S*s in Group 1 ran significantly faster after 1- than after 10-sec. delays in the postshift phase, $F(1, 7) = 6.72$, $p < .05$. Except for this single significant effect, which agrees with the data of Wist (1962) but contradicts that of Holder et al. (1957), this study provides no evidence that postdelay response strength is a function of different amounts of delay beyond a zero value.

DISCUSSION

The principal implications of the present experiment stem from gross changes in response strength over trials. The pronounced decline in A_1 speeds following an increase in the first 25–36 trials, together with the sharp rise in the incidence of retracing, clearly supports the conclusion that at least one of the effects of a temporal delay of reinforcement is an inhibitory or subtractive effect on reaction potential. These results also supplement those of Logan (1960) who found, under somewhat different experimental conditions, that beginning around Trials 15–30, speeds for some delay groups began to decrease, albeit less dramatically than in the present study. In addition, he reports that the decremental effect appears to move backward through the alley with training. The present data also substantiate this finding as may be seen by comparing Fig. 1 and 2.

If it is assumed that delay produces a hypothetical emotional response of frustration, then the results can be explained by the theoretical assumptions of Amsel (1958) and Spence (1960). Within the theory anticipatory frustration, r_f, the fractional conditioned form of primary frustration, becomes elicited by alley cues antedating the aversive delay. The decremental effects on speed, then, are attributable to competing or avoidance responses mediationally aroused by the characteristic cues, s_f, associated with r_f. Furthermore, these decremental effects do not occur earlier in training because there is no frustration aroused until anticipatory reward, r_g, has had opportunity to reach some minimal strength.

A second, particularly interesting, finding of the present study is the recovery of A_1 speeds late in training and the correlated decrease in A_2 starting speeds. Similar to the onset of the decremental effect, recovery of A_1 speeds may begin at different points in training depending on the segment of the response chain measured. Figures 1 and 2 suggest that the recovery first appears in the running speeds and then in the starting speeds.

Perhaps the most obvious interpretation of this second set of results is that both primary frustration and anticipatory frustration were attenuated late in training. Speeds in A_1 recovered because r_f, which mediated the decremental effect, disappeared, and speeds in A_2 decreased because the increment to drive from primary frustration was removed.

Why should the delay chamber lose its aversive properties with the result that both avoidance of it and escape from it are diminished? A probable explanation is that r_g, which originally generalizes from the goal box, eventually undergoes extinction in A_1. Within frustration theory if r_g disappears so should r_f. In effect, *S*s may learn via the differential extinction of r_g, to discriminate A_1 leading to a delay from A_2 which leads to a reinforcement. This assumption is made plausible by recent studies (Ward, cited by Amsel, 1962; McHose, 1963) employing double-runway apparatus, which indicate that the facilitative effects of frustrative nonreward dissipate when anticipation of reward extinguishes.

The hypothesis presented here may be applied to certain groups from double-alley studies (McHose, 1963; Wagner, 1959) which show results quite unlike those obtained in the present study despite similar delay procedures. The pertinent data from these studies may be summarized by stating that there was no pronounced decrement and recovery in A_1 speeds, nor was there a decrement in A_1 speeds late in training. According to the hypothesis of this paper the discrepancy in results is attributable to the fact that A_1 and A_2 were highly dissimilar in the Wagner and McHose studies but quite similar in the present study. Thus only in the latter case would r_g be expected to generalize substantially to A_1 and thereby produce the observed frustrative effects.

The problem of the mechanism underlying effects of different amounts of delay beyond a zero value remains. The present study clearly supports the notion that frustration is, at least at times, a factor in delay of reward, but it is not clear whether the degree of frustration is a function of the length of delay. Neither does it seem possible that frustration-mediated avoidance can account for all the decremental effects of delay in view of the early onset of such effects in a number of studies. Although presumably not mediated by r_f, such early effects might still result, in the manner suggested by Spence (1956), from interfering responses which have occurred during the delay and become conditioned to essentially the same stimulus components which elicit the instrumental response. In any case, it seems likely that delay can no longer be considered a unitary factor in behavior, but that several processes—some of which may change with learning—will have to be postulated.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychol. Rev.*, 1962, **69**, 306–328.

HOLDER, W. B., MARX, M. H., HOLDER, E. E., & COLLIER, G. Response strength as a function of delay of reward in a runway. *J. exp. Psychol.*, 1957, **53**, 316–323.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

LOGAN, F. A. The role of delay of reinforcement in determining reaction potential. *J. exp. Psychol.*, 1952, **43**, 393–399.

LOGAN, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

MCHOSE, J. H. The effect of continued nonreinforcement on the frustration effect. *J. exp. Psychol.*, 1963, **65**, 444–450.

PERIN, C. T. A quantitative investigation of the delay-of-reinforcement gradient. *J. exp. Psychol.*, 1943, **32**, 37–51.

RAMOND, C. K. Performance in instrumental learning as a joint function of delay of reinforcement and time of deprivation. *J. exp. Psychol.*, 1954, **47**, 248–250.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs: Prentice-Hall, 1960.

WAGNER, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *J. exp. Psychol.*, 1959, **57**, 130–136.

WIST, E. R. Amount, delay, and position of delay of reinforcement as parameters of runway performance. *J. exp. Psychol.*, 1962, **63**, 160–166.

(Received June 24, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 412–416

TEMPORAL DURATION: RATIO SCALE AND CATEGORY SCALE[1]

RAM G. CHATTERJEA[2]

Brown University

3 groups of Ss judged 2-, 4-, 8-, 16-, and 32-sec. durations of time presented through flash of light by fractionation, multiplication, and category methods. The stimuli were controlled by a Hunter timer and the ratio production judgments recorded by a Standard Electric .01-sec. clock. Apparent duration appeared to grow as the 1.02 power of stimulus magnitude while the category scale was nonlinear. Discrepancies in half and double judgments were found which were ascribed to the underestimating bias throughout the stimulus range.

Scaling of a psychological magnitude for time has been attempted by the method of ratio production (Ekman & Frankenhaeuser, 1957; Gregg, 1951; Mallick, 1962; Ross & Katchmar, 1951). The first three studies used auditory stimuli to mark the intervals while Mallick (1962) used a visual one. Although intervals ranged from 400 msec. to 60.13 sec., there is agreement in the results that subjective time is a power function of physical time with an exponent close to 1.00. In contrast to linearity of a ratio scale, one would expect a curvilinear function from a category scale (Stevens & Galanter, 1957).

The purpose of the present study was to determine the stability of the ratio scale by comparing the results of halving and doubling and to see how this scale was related to one determined by category judgments.

METHOD

Subjects.—The *S*s were 40 graduate students at Brown University. Nine men and 5 women with a mean age of 22.8 yr. (range 17–28 yr.) served in the fractionation experiment. The multiplication-estimation experiment employed 9 men and 6 women having the same mean age and range as that of the previous one. In the category judgment there were 7 men and 4 women with a mean age of 26.2 yr. (range 19–45 yr.).

Procedure.—A Hunter timer was used to control the stimulus durations which were indicated by flashes with a 2.5-w. bulb placed in front of *S*. The durations consisted of 2, 4, 8, 16, and 32 sec. The *S*s responded by pressing a key and their judgments were recorded by a Standard Electric .01-sec. clock.

The instructions called upon *S*s to refrain from keeping track of the duration either by counting or making rhythmical movements. In the halving experiment *S*s were told to depress the key for a duration just half that of the flash, while in the doubling experiment they were to press the key for a period twice the duration of the flash. The instruction for the category judgment was as follows:

> I will present to you one of several durations by lighting the bulb. I want you to assign a number, 1, to the smallest duration through 5 for the longest one. Before you actually see the flashes in my experimental series, I will show you one, as this one, that is slightly shorter than the shortest in the series and another which is slightly longer than the longest in my experimental series.

The durations were presented in an irregular order to each *S* and each stimulus duration was presented three times.

RESULTS

The mean of three judgments for each stimulus time was determined. The median for the group was then

---

[1] The study was supported in part by National Science Foundation Grant GB-724.

[2] The author is a member of the staff of the University Psychology Laboratory, Calcutta University, India.

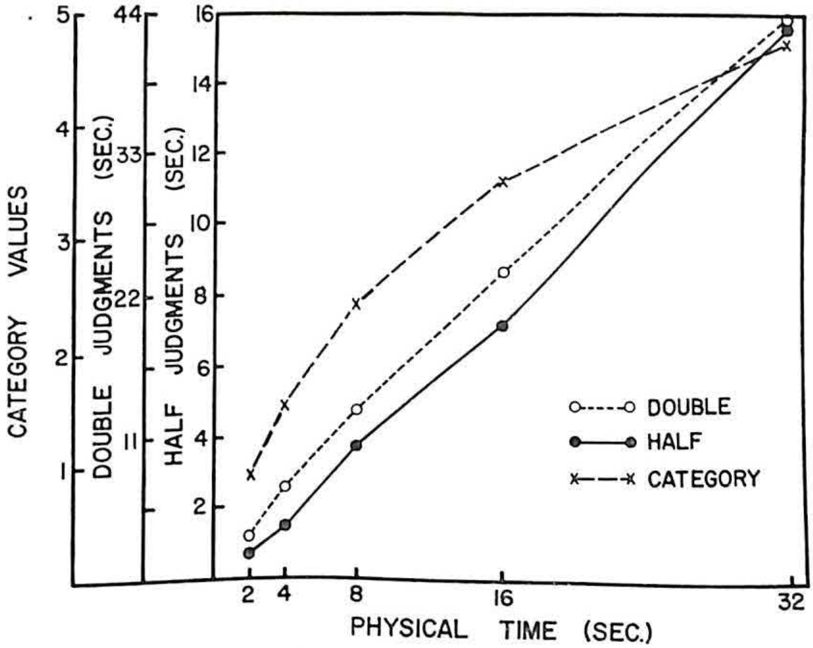


FIG. 1. Ratio and category judgments plotted in linear coordinates. (Closed circles: median judgments of 14 *Ss*; open circles: median judgments of 15 *Ss*; dotted line: median category judgments made by 11 *Ss* on a 5-point scale from 1 to 5.)

taken for final analysis. A plot of the half and double judgments as well as the category values was made (Fig. 1). From the half-judgment curve the psychological scale termed *chron* was obtained following the precedure outlined by Harper and Stevens (1948). One *chron* is thus the subjective time

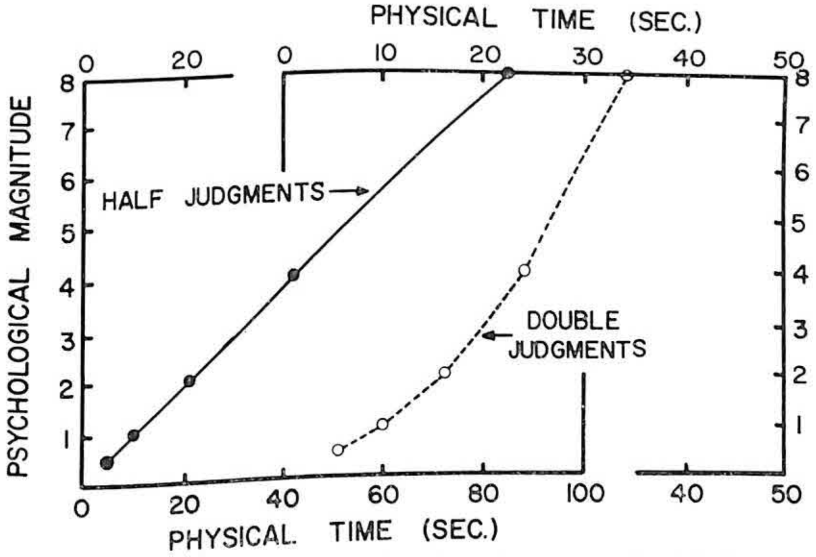


FIG. 2. Psychological magnitude plotted against physical magnitude in linear coordinates. (Closed circles: chron scale obtained from half judgments; open circles: chron scale derived from double judgments.)

corresponding to physical time of 10 sec. Using the same method as that of Ross and Katchmar (1951), we assigned D to a given magnitude and D/2 to the one judged half and constructed the chron scale from the half-judgment curve of Fig. 1. Since a physical time of 4.8 sec. was judged half as long as that of 10 sec., we assigned .5 (subjective time) to 4.8 sec. and 1 (subjective time) to the physical time of 10 sec. We could have similarly found physical times corresponding to psychological times of 2, 4 chrons, etc. Figure 2 shows the physical times in seconds corresponding to subjective times in chrons.

Using a similar procedure, we were able to construct another scale from the double-judgment curve of Fig. 1. Since a physical time of 15.4 sec. was found to be judged to be twice as long as that of 10 sec., we assigned 2 (subjective time) to the value of 15.4 sec. and 1 (subjective time) to the physical time of 10 sec. This scale is also shown in Fig. 2.

The fact that the nature of the data falls close to a straight line on a log-log plot indicates that apparent duration in this case is a power function of stimulus duration. The least-square solution for the half- and double-judgment data of Fig. 1 yields the following exponents: $C = .19D^{1.02}$ for $\frac{1}{2}$ judgment and $C = .87D^{.78}$ for double judgment, where C is subjective time in chrons and D is stimulus duration in seconds (Lewis, 1960).

The discrepancy in the two exponents is due to the underestimating bias irrespective of stimulus magnitude which will be evident from Fig. 3. The proportions of underestimation are clear in all the cases, although in double judgments they are conspicuously great. The category scale is nonlinear to the ratio scale in this case, too. When plotted against the ratio judgments (Fig. 4), the trend is, however, the same as is found in other prothetic continua (Stevens, 1960; Stevens & Galanter, 1957).

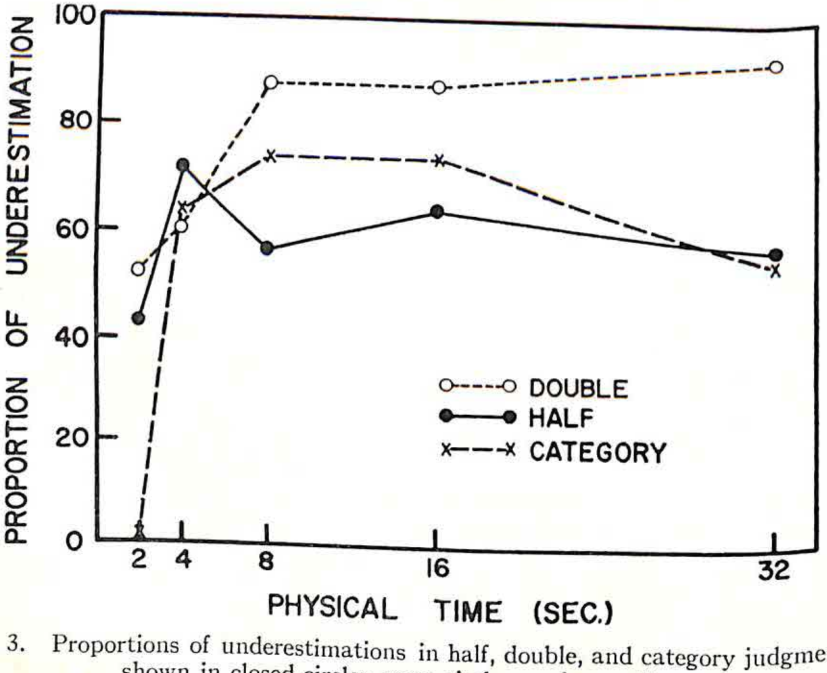


FIG. 3. Proportions of underestimations in half, double, and category judgments, shown in closed circles, open circles, and cross lines.

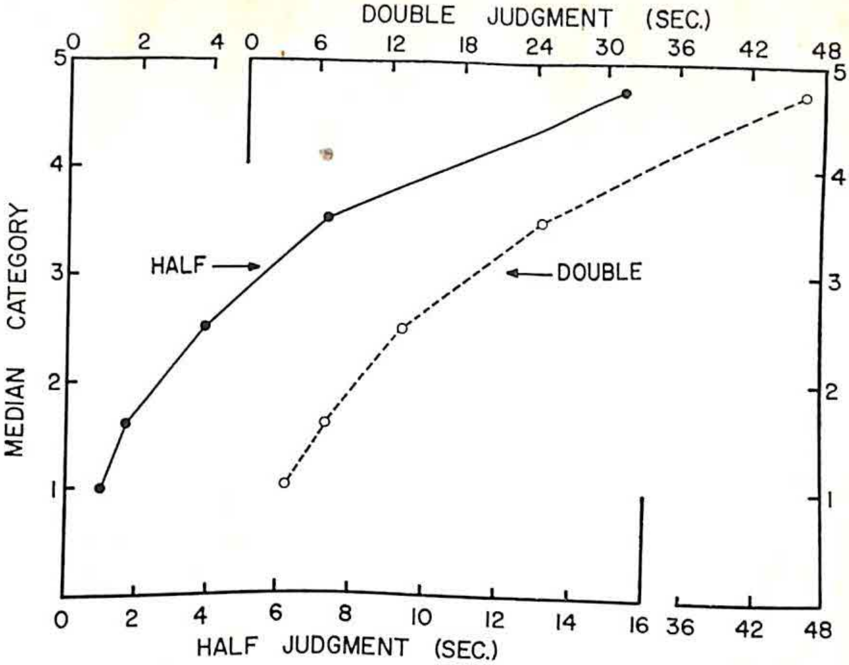


FIG. 4. Category scale plotted against ratio judgments. (Closed circles show half while open circles denote double judgments.)

DISCUSSION

In conformity with other perceptual continua, the fractionation data in this study reveal that the psychophysical relation for duration is a power function with an exponent of 1.02. The exponent in case of doubling data turns out to be .78. It appears that effects other than operant behavior on the part of *S*s might have influenced doubling judgments. If we assume that halving and doubling are subject to a bias that operates in an opposite direction (Stevens, 1955, p. 821), we would expect a counterbalancing in the two judgments. On this analogy then, judging half of 32 sec. would be the reverse of judging double of 16 sec. But the results do not agree with this view. And why?

One possible explanation seems to be that *S*s find it somewhat boring to wait and therefore terminate their judgments too soon. This is especially true for stimulus durations of longer magnitudes covered in this study. In other words there is a consistent bias for underestimation in case of doubling judgments as is apparent from Fig. 3. The proportion of underestimation is well above 80% in three of the five stimulus magnitudes, while halving judgments show much less at these places. However there are also effects due to spacing of the stimuli—effects which need further study before we could attempt any generalization.

The category scale is clearly nonlinear which is in line with other studies. Though the stimuli were not approximately equally spaced in this study the curve is concave downward (cf. Stevens & Galanter, 1957, pp. 380–381). When plotted against half and double judgments the category scale turns out to be concave downward (Fig. 4). This is also in line with studies reported for other perceptual continua (Stevens, 1958).

REFERENCES

EKMAN, G., & FRANKENHAEUSER, M. Subjective time scales. Report No. 49, 1957, University of Stockholm, Psychological Laboratory.

GREGG, L. W. Fractionation of temporal intervals. *J. exp. Psychol.*, 1951, **42**, 307–312.

HARPER, R. S., & STEVENS, S. S. A psychological scale of weight and a formula for its derivation. *Amer. J. Psychol.*, 1948, **61**, 343–351.

LEWIS, D. *Quantitative methods in psychology.* New York: McGraw-Hill, 1960.

MALLICK, A. K. A scale for the measurement of subjective duration. *J. psychol. Res., Madras*, 1962, **6**, 1–3.

ROSS, S., & KATCHMAR, L. The construction of a magnitude function for short time intervals. *Amer. J. Psychol.*, 1951, **64**, 397–401.

STEVENS, S. S. The measurement of loudness. *J. Acoust. Soc. Amer.*, 1955, **27**, 815–829.

STEVENS, S. S. Problems and methods of psychophysics. *Psychol. Bull.*, 1958, **54**, 177–196.

STEVENS, S. S. Ratio scales, partition scales and confusion scales. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling.* New York: Wiley, 1960.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

(Received May 8, 1963)

ERRATUM

In the supplementary report by Dallett (*J. exp. Psychol.*, 1964, **67**, 296–298) the M value for Cond. 07:07 in Table 1 (p. 297) should read 4.79.

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 417–422

MEAN AMOUNT OF REINFORCEMENT AND INSTRUMENTAL RESPONSE STRENGTH [1]

STEWART H. HULSE AND ROBERT J. FIRESTONE

Johns Hopkins University

Licking was treated as a free operant using a mean amount of reinforcement per lick of .0010 ml. of a saccharine solution. The fixed mean amount of reinforcement was determined by combining appropriate absolute volumes of reinforcement with 3 fixed ratios of reinforcement: continuous reinforcement (CRF), 2 to 1 fixed ratio (FR 2), and 8 to 1 fixed ratio (FR 8). 3 training methods varied the way in which Ss were shaped toward their final common mean amount of reinforcement. Acquisition rates of licking were essentially identical for all groups, while the number of licks emitted during extinction increased as ratio of reinforcement increased. The results imply that acquisition response rates are constant for a given mean amount of reinforcement, while resistance to extinction is positively correlated with the variability of absolute amounts that S obtains from response to response during acquisition.

Hulse, Snyder, and Bacon (1960) suggested that the mean amount of reinforcement that an *S* receives per response during instrumental conditioning is an important parameter of reinforcement in learning situations involving partial reinforcement. They noted that an *S* that is on a 5 to 1 fixed ratio (FR 5) receives a mean amount of reinforcement per response that is one-fifth that obtained by an *S* on continuous reinforcement (CRF) when the absolute physical amount of the reinforcing stimulus—when it is delivered—is the same for both *S*s. In an experiment which combined two absolute volumes of a fluid reinforcing stimulus with CRF and FR 5, their data were most meaningfully described by mean volume of reinforcement. As mean volumes increased, the rate of an instrumentally conditioned licking response increased, but the number of licks that was emitted during extinction decreased. The present experiment was designed to explore further the utility of the concept of mean amount of reinforcement in instrumental conditioning.

There are three ways in which mean amount of reinforcement may be varied in conditioning with partial reinforcement. First, ratio of reinforcement can be held constant while the absolute amount of the reinforcing stimulus is varied. Second, absolute amounts of the reinforcing stimulus can be held constant while ratio of reinforcement is varied. Third, mean amount of reinforcement can be set as an independent variable and different mean amounts obtained by simultaneously varying *both* ratio of reinforcement and absolute amount of reinforcement. In the first two instances, which are characteristic of all experiments that we know of which have studied amount and ratio of reinforcement, mean amount of reinforcement is perfectly confounded with amount of reinforcement or with ratio of reinforcement. In the typical partial reinforcement experiment, for example, the amount of the reinforcing stimulus is fixed while ratios of reinforcement are varied. As ratio of reinforcement increases, the mean amount of reinforcement obtained with each response necessarily decreases. In the third instance, where mean amount of reinforcement is chosen as the independent variable, a particular mean amount and a particular ratio of reinforcement can be set freely, but the absolute amount of reinforcement that must be used is then determined. For example, choose 10 units of reinforcement as a mean amount of reinforcement per response, and FR 2 and FR 5 as ratios of reinforcement. For the FR 2 condition, *S* would actually receive 20 units of reinforcement each time the reinforcing stimulus was delivered. For FR 5, the appropriate value would be 50 units of reinforcement. Conversely, of course, a mean amount and an absolute amount could be set, in which case a particular ratio of reinforcement would be determined.

---

[1] This paper, which is based on the master's essay of the second author, was supported by Grant G18125 from the National Science Foundation to the first author.

How do organisms respond to mean amounts of reinforcement as defined above? If mean amount of reinforcement is an important parameter of the strength of a response, it is possible that equal mean amounts of reinforcement would produce equal response strengths. This would follow from the classic notion in learning theory that the strength of a response is correlated with the amount of reinforcement that is used to reward the response, and it would also follow from the Hulse et al. (1960) data cited above. The research reported here was designed with this point of view in mind.

The experiment incorporated the instrumentally conditioned tongue lick in a free responding situation. Mean amount of reinforcement was set at .0010 ml. of a saccharine solution, and three fixed ratios of reinforcement were used: CRF, FR 2, and FR 8. In order to satisfy the mean amount criterion, the absolute volumes of fluid per reinforcement for the three ratios of reinforcement varied from .0010 ml. to .0080 ml. Three training procedures were used which varied the extent to which mean volume of reinforcement was confounded with ratio of reinforcement as *S*s in the partial groups were shaped toward their final schedules of reinforcement.

METHOD

Subjects and Apparatus

The *S*s were 49 experimentally naive, male albino rats of the Sprague-Dawley strain obtained from Sprague-Dawley, Incorporated, Madison, Wisconsin. The *S*s were approximately 80 days of age at the time the experiment began, and none had had experience drinking from tubes prior to their use in the experiment.

The apparatus has been described in detail elsewhere (Hulse, 1960; Hulse et al., 1960). Briefly, it consisted of a lighted, sound-shielded wooden box 6.5 × 10 × 7 in., with a Plexiglas top. One wall, made of Plexiglas, had a 1 × 2 cm. vertical slot cut into it. The *S* had access to a drinking tube through the slot.

A liquid reinforcement system was used which made obtaining a drop of fluid from a drinking tube contingent on a tongue lick. The system included an electronic relay (Otis & Boenning, 1959) which operated a programing circuit each time *S*'s tongue touched a metal contact on the drinking tube. The programmer, in turn, operated an infusion pump which delivered a drop of fluid of known size to the tip of the drinking tube through a piece of pressure tubing. Two drinking tubes were used: a brass training tube, and a Plexiglas tube. Both tubes had 2-mm. fluid holes in their centers. The Plexiglas tube had a $\frac{1}{16}$-in. brass contact located $\frac{1}{16}$ in. below the fluid hole which provided a contact for the electronic relay. A solenoid-driven lever produced a slight negative pressure in the fluid system on nonreinforced licks, assuring that *S*'s tongue could not touch the fluid column on nonreinforced licks.

Procedure

Experimental design.—Three fixed ratios of reinforcement were used: CRF, FR 2, and

FR 8. A terminal volume for the reinforcing stimulus was chosen such that at the end of acquisition mean volume of reinforcement per lick was essentially identical for all ratios of reinforcement (.0010 ± .0001 ml.). The terminal volumes were .0010, .0020, and .0080 ml. for CRF, FR 2, and FR 8, respectively. For the FR 2 and FR 8 ratios, the mean volume of .0010 ml. was approached according to one of three methods. With a constant-volume method (CM), *S*s in the partial groups were placed directly on their terminal volumes at the beginning of acquisition. Acquisition began on CRF, and ratios of reinforcement were increased to their final levels as acquisition progressed. With this method, *S*s who were to be on FR 8 began training with a larger mean volume of reinforcement per response than *S*s who were to be on FR 2. With an adjusted-volume method (AM), *S*s in the FR 2 and FR 8 groups began training on CRF with the same volume of the reinforcing stimulus (.0010 ml.), but the volume of the reinforcing stimulus was increased as *S*s were shaped toward their final ratio of reinforcement. With this method, the mean volume of reinforcement per lick was constant at .0010 ml. throughout all stages of training. With an initial method (IM), an attempt was made to place *S*s directly on their final FR 2 and FR 8 ratios at the beginning of training, with volume of the reinforcing stimulus set at .0020 and .0080 ml. Thus, with respect to the FR 2 and the FR 8 ratios of reinforcement, and with respect to the three methods of training, the design was a simple factorial involving six independent groups of *S*s. A seventh group was run on CRF with absolute volume of reinforcement—hence mean volume of reinforcement per response—set at .0010 ml. Seven *S*s were randomly assigned to each of the experimental groups.

Deprivation and taming.—Upon receipt from the breeder, *S*s were placed in colony cages and maintained on ad-lib. Purina laboratory chow and water (available from cans) for 2 days. They were then transferred to individual cages without food or water for 1 day. This was followed by 13 days of taming. Each day *S*s were freely handled and allowed to explore in groups in large wooden boxes for 20–30 min.

On Day 1 of taming, each *S* was placed on a daily deprivation regimen of 9 gm. of ground Purina chow mixed with 15 ml. of tap water. This schedule was maintained for the duration of the experiment. The *S*s were fed the wet mash such that they were approximately 23 hr. hungry at the time the experimental procedure began each day.

Acquisition and extinction of licking.—Following taming, *S*s received 33 days of acquisition of the instrumental licking response. Each *S* was permitted 400 licks per day, or 3 min. in the experimental box, whichever occurred first. A fluid consisting of 10 gm. saccharine added to 1 l. of tap water was used as the reinforcing stimulus.

On Day 1, *S*s licked the brass training tube which extended through the slot $\frac{1}{4}$ in. into the experimental box. This tube was gradually withdrawn so that by Day 6 it was centered $\frac{1}{16}$ in. behind the slot. On Day 7, the brass tube was replaced with the Plexiglas tube, and the latter was used for the remainder of the experiment.

On Days 1 through 10, *S*s in the CRF and in the CM and AM groups licked on a CRF schedule of reinforcement with the volume of the reinforcing stimulus set according to the experimental design. On Days 11 through 17, *S*s in the CM and AM groups were shaped toward their final schedules, while adjustments were made in the volume of the reinforcing stimulus for the AM group. All *S*s then licked on their final ratios and volumes for the remaining 16 days of acquisition.

Following the last day of acquisition, 4 days of extinction began. The procedure was exactly the same as during the last 16 days of acquisition, except that the pump system was mechanically disengaged so that it could not deliver fluid. On each extinction day, *S*s were permitted 3 min. in the box during which time they could lick the dry tube.

Measures of Performance

During acquisition, the time that *S* required to emit its 400-lick allotment was recorded each day. These times were converted to licking rates in licks per second. Licking rates were averaged over the last 3 days of acquisition for each *S* for purposes of the statistical analyses.

During extinction, the number of licks that each *S* emitted at the dry tube was recorded for each of the four extinction sessions. Number of licks was used for statistical analysis of extinction performance.

RESULTS

Acquisition.—Inspection of the average rates of licking for the last 3 days of acquisition (Table 1) suggests no consistent trend in the data correlated with either ratio of reinforcement or training method. Moreover, all the

TABLE 1

MEANS AND *SD*s FOR LICKING RATES DURING ACQUISITION AND FOR NUMBER OF LICKS DURING EXTINCTION

Group	Acquisition		Extinction	
	M	*SD*	*M*	*SD*
CRF	4.05	0.58	325	110
FR 2 AM	3.97	0.31	344	152
FR 2 CM	3.89	0.30	321	186
FR 2 IM	3.90	0.43	430	309
FR 8 AM	3.88	0.40	428	394
FR 8 CM	4.02	0.40	423	200
FR 8 IM	3.92	0.20	455	110

mean rates of licking are within approximately 0.1 lick per second of each other.

Statistical analysis of the acquisition data was carried out in three ways. First, an overall analysis of variance was done on the data for all seven groups combined according to ratio of reinforcement. Second, three separate analyses of variance were done, each of which compared the CRF group with the FR 2 and the FR 8 group run under a particular method of training. Third, the CRF condition was disregarded and an analysis of variance was done on the FR 2 and FR 8 groups conditioned with the three methods of training. The results of the first analysis revealed a nonsignificant effect ($p > .05$) for ratio of reinforcement. The results of the second set of analyses revealed nonsignificant effects for ratio of reinforcement for each of the three methods of training. The results of the third analysis showed nonsignificant effects for ratio of reinforcement, method of training, and the interaction between these two variables.

Extinction.—The number of licks emitted on each extinction day by each *S* (Table 1) was subjected to the same pattern of analysis that was used on the acquisition data. The analysis done on the seven groups combined according to ratio of reinforcement showed this variable to produce a nonsignificant effect, although the table shows a clear trend for resistance to extinction to increase as ratio of reinforcement increases. Similarly, when the CRF group was compared with the FR 2 and the FR 8 group run with a particular method of training, nonsignificant effects for ratio of reinforcement were also obtained. The analysis which examined the effects due to the FR 2 and the FR 8 ratios, and the three methods of training, showed nonsignificant effects for ratio of reinforcement and method of training. However, this analysis did reveal a significant interaction between ratio of reinforcement and days of extinction, $F(3,108) = 2.78, p < .05$. A *t* test showed that the FR 8 group was more resistant to extinction than the FR 2 group on Day 1 of extinction, $t\ (108) = 3.83,\ p < .01$. On subsequent days of extinction, however, the difference in number of licks emitted by these two groups was not significant.

DISCUSSION

There are two straightforward conclusions to be drawn from the data. First, although there is risk in accepting no-difference data as final, the acquisition results argue strongly that equal mean volumes of reinforcement produce equal rates of responding under the conditions of the present experiment. With mean volume held constant, rate of responding does not appear to change with the ratio of reinforced to nonreinforced responses per se, nor with the absolute volume of the reinforcing stimulus per se.[2] These data are completely consistent with acquisition data obtained for rats in a straight runway (Logan, Beier, & Kincaid, 1956) and are congruent with data obtained for human maze performance (Bevan & Adamson, 1960). Second, when mean volume of reinforcement is held constant during acquisition, different ratios of reinforcement still produce a characteristic partial reinforcement effect during extinction. As ratio of reinforcement increases, so does resistance to extinction, although comparison of the data with some data obtained earlier suggests that the partial reinforcement effect may be reduced to some extent. Thus, in an experiment which was essentially identical in details of procedure to this one—except that mean volume of reinforcement was confounded with ratio of reinforcement—Hulse and Bacon (1962) obtained a much larger partial reinforcement effect for FR 8 as compared with CRF.

---

[2] This argument can be clinched by an experiment, currently under way, which incorporates more than one mean volume of reinforcement. For a given mean volume, equal rates of responding should be obtained across ratios or absolute volumes of reinforce-

The results can be accounted for by means of a model which is quite similar to one suggested by Logan (1960). During acquisition, a given mean amount of reinforcement per response is set. However, the absolute physical amounts of reinforcement actually given *S* from response to response are systematically varied according to some preset distribution of amounts about the mean amount. Continuous reinforcement becomes a special case where, for a given mean amount, the variance of the amounts actually given *S* is 0, i.e., *S* receives the same amount following each response. Partial reinforcement becomes a special case where, for a given mean amount, the distribution of amounts given *S* is chosen to include a 0 amount, i.e., nothing at all, following some responses. There are, of course, an infinite variety of statistical distributions of amounts from which partial reinforcement could be derived as a special case.

The available data suggest that, within general limits, response strength during acquisition is positively correlated with the mean of the distribution of amounts. For a given mean amount, however, resistance to extinction is positively correlated with the variance, or variability, of the distribution of amounts that *S* receives. Further, some data of Hulse (1962), and some data obtained in an ingenious experiment by Berkley (1963) suggest that variable amounts increase resistance to extinction to the extent that the different amounts in a distribution are discriminable from one another as stimuli. Berkley conditioned the licking response by presenting a drop of fluid every tenth lick that was constant in both volume and concentration of saccharine. On the intervening nine licks, *S* received reinforcing stimuli which varied along two dimensions: volume, and concentration of saccharine. Volumes were chosen to include nothing at all, and a volume that was the same as that presented every tenth lick, as special cases (partial and continuous reinforcement with respect to volume). Concentrations were chosen to include water, and a concentration that was the same as that presented every tenth lick, as special cases (partial and continuous reinforcement with respect to sweetness). In general, resistance to extinction increased as the contrast increased in the qualitative and quantitative properties of the reinforcing stimuli presented from trial to trial during acquisition.

We should like to conclude by pointing explicitly to a paradox that lurks in the data that are obtained with free operant licking as compared to the data that are obtained with other kinds of free operant responses. The paradox has potential implications for the usefulness of mean amount of reinforcement as a meaningful parameter of free operant behavior. If rats or pigeons are turned loose in conventional experimental spaces, rates of responding increase dramatically when ratios of reinforcement increase. When licking is treated as a free operant, however, rates of responding *decrease* dramatically when ratios of reinforcement increase (Hulse & Bacon, 1962; Hulse et al., 1960). As we have seen, mean amount of reinforcement necessarily be-

---

ment, but as mean volume increases, progressively higher rates of responding should be obtained.

comes smaller as ratios of reinforcement increase so long as the absolute amount of reinforcement is held constant. Therefore, for most free operants, rate of responding increases as mean amount of reinforcement becomes *smaller*. For licking, however, the converse is true. The potential solution to this paradox lies in at least two possible directions. First, experiments incorporating the present approach simply have not been carried out with free operant responses other than licking. If mean amount were set as a parameter, and absolute amounts varied from bar press to bar press, for example, data more closely analogous to that obtained here might be found. In this connection, it may be that rate of responding as a measure of response strength for the variables under discussion is not appropriate in situations where *S* has to move from the manipulandum to the reinforcement magazine to eat or drink. When amount of reinforcement is varied—particularly for low ratios of reinforcement—eating or drinking time becomes an increasingly large component of the times that are used to compute response rate on the bar. With licking, this problem is avoided since the response that produces the reinforcer is at the same time the response that consumes the reinforcer. Finally, it is possible that the behavior that is characteristic of the licking response is unique. There are some suggestions in the literature (Stellar & Hill, 1952) that the tongue lick has many of the properties of a fairly low-order reflex. Perhaps the function of the reinforcing stimulus in free operant licking is to *elicit* tongue licks, as much as it is to reinforce tongue licks after they are emitted.

REFERENCES

Berkley, M. A. Discrimination of rewards as a function of contrast in reward stimuli. *J. exp. Psychol.*, 1963, **66**, 371–376.

Bevan, W., & Adamson, R. Reinforcers and reinforcement: Their relation to maze performance. *J. exp. Psychol.*, 1960, **59**, 226–232.

Hulse, S. H. A precision liquid feeding system controlled by licking behavior. *J. exp. Anal. Behav.*, 1960, **3**, 1–3.

Hulse, S. H. Discrimination of the reward in learning with partial reinforcement. *J. exp. Psychol.*, 1962, **64**, 227–233.

Hulse, S. H., & Bacon, W. E. Supplementary report: Partial reinforcement and amount of reinforcement as determinants of instrumental licking rates. *J. exp. Psychol.*, 1962, **63**, 214–215.

Hulse, S. H., Snyder, H. L., & Bacon, W. E. Instrumental licking behavior as a function of schedule, volume, and concentration of a saccharine reinforcer. *J. exp. Psychol.*, 1960, **60**, 359–364.

Logan, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

Logan, F. A., Beier, E. M., & Kincaid, W. D. Extinction following partial and varied reinforcement. *J. exp. Psychol.*, 1956, **52**, 65–70.

Otis, L. S., & Boenning, R. A. A transistorized circuit for recording contact responses. *J. exp. Anal. Behav.*, 1959, **2**, 289–291.

Stellar, E., & Hill, J. H. The rat's rate of drinking as a function of water deprivation. *J. comp. physiol. Psychol.*, 1952, **45**, 96–102.

(Received June 3, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 423–431

DRIVE AND INCENTIVE VARIABLES ASSOCIATED WITH THE STATISTICAL PROPERTIES OF SEQUENCES OF STIMULI[1]

AUSTIN JONES

University of Pittsburgh

Drive and incentive variables associated with the information, complexity, and fluctuation of series of stimuli were compared. Visually deprived human Ss could execute an instrumental response which introduced series of light flashes varying in their degree of information, complexity, and fluctuation. A drive interpretation of information deprivation was supported by increasing response rates over hours of deprivation ($p < .01$); a drive interpretation was not supported for complexity or fluctuation deprivation. Response was found to be an increasing function of the relative value of all 3 incentive dimensions (p's $< .001$, $<.05$, $<.05$, respectively). The information-incentive function was superordinate to the other 2 in that they were suppressed when information incentives were simultaneously available to Ss.

Recent studies have suggested that visual information deprivation in humans generates a drive and that visual information functions as an incentive (Jones, 1961; Jones, Wilkinson, & Braden, 1961). The *S*s maintained in light-proof chambers were permitted instrumental responses serving to introduce patterned series of red and green lights. Response rate was an increasing function of hours of deprivation and of the information (uncertainty) value of the series. The purpose of the present paper is to report a series of experiments which provides a comparison of the drive and incentive properties associated with stimulus information with those of two additional statistical properties of stimulus series, *complexity* and *fluctuation*. The drive-incentive relationships associated with these dimensions are evaluated both individually and in "competition" with each other and with the information dimension, so that the hierarchical organization or relative dominance of the three dimensions may be established.

Stimulus *information* refers to the degree to which the succession of stimuli is unpredictable by *S*. Stimulus *complexity* is defined as the degree of "physical" randomness in a series. A stimulus series may have a high degree of complexity but zero information, as when a particular randomly ordered sequence, repeatedly presented to *S*, becomes predictable by him. Stimulus *fluctuation* refers to the degree to which stimulus categories alternate. A stimulus series may have a high degree of complexity but zero information, as, for example, the single alternation A, B, A, B. . . . There is, then a logical hierarchical organization of the dimensions in the order information, complexity, fluctuation. Nonzero values of each dimension are always accompanied by nonzero values of dimension(s) subordinate to it, while superordinate dimensions may have zero values. A zero value of a dimension requires zero values for superordinate dimensions, while sub-

---

[1] This study was supported in part by Grant M-2479 from the National Institute of Mental Health, United States Public Health Service. The author is indebted to Peter Bentler for assistance in the execution of the experiments and to Harry Fowler for comments on the manuscript.

TABLE 1

SAMPLE VALUES OF THE INFORMATION, COMPLEXITY, AND FLUCTUATION VARIABLES

Relative Value	Information	Complexity (Information at .00)	Fluctuation (Information and Complexity at .00)
	Serial Order of Flash: 1 2 3 4 . . .	Serial Order of Flash: 1 2 3 4 . . .	Serial Order of Flash: 1 2 3 4 . . .
1.0	R R R R . . .	r r r r . . .	a b a b . . .
.5	a R a R . . . or b R b R . . .	a r a r . . . or b r b r . . .	a a b b a a b b . . . or b b a a b b a a . . .
.0	a a a a . . . or b b b b . . .	a a a a . . . or b b b b . . .	a a a a . . . or b b b b . . .

Note.—R is a flash whose color is randomly determined anew on each selection of the series by *S*; r is a flash whose color is initially randomly determined but is thereafter fixed upon further selections of the series; a and b represent red and green flashes whose sequential position is either repetitive or symmetrically alternating and is invariant over repeated selections of the series.

ordinate dimensions may be nonzero. Thus a sequence cannot contain information without containing also some degree of complexity and fluctuation, and a series cannot contain complexity without also some degree of fluctuation. In the experiments which follow, stimulus series of red and green light flashes embodying varying degrees of these dimensions were constructed in the manner outlined in Table 1.

EXPERIMENT I

In the experiments by Jones et al. (1961), response rate increased monotonically as a function of the information value of the light series. That the crucial aspect of the light incentives was the information dimension is not entirely clear, however, for the reason that increased values of information (uncertainty) of the light series, were paralleled by and confounded with increasing values of complexity ("physical" randomness) of the light series. The purpose of the first of the present experiments was to provide a comparison of the relative strengths of the drive-incentive relationships associated with these two dimensions. The comparison was a simultaneous one, in which *S*s deprived of both stimulus information and complexity were permitted to make responses serving to introduce both information and complexity incentives. Complexity series, throughout this paper, will refer to series having some nonzero value of complexity but with zero information value. As noted earlier, information series necessarily contain complexity values as well, since a series cannot be uncertain or unpredictable without also having some degree of physical randomness. Thus, in the present experiment, should response be shown to have the same linear function over both information and complexity values, then the crucial dimension of the incentive would be established as complexity, since the complexity series contain no information.

The simultaneous visual conditions of information and complexity deprivation were provided by isolating *S*s in light-proof chambers as described in the earlier experiments (Jones et al., 1961). The *S*s were permitted a selection of light series among comparable ranges of information and complexity values.

The hypotheses concerned with drive interpretations were as follows: (*a*) The rate of instrumental response for information series will be an increasing monotonic function of hours of deprivation. (*b*) The rate of instrumental response for complexity series will be an increasing monotonic function of hours of deprivation. (*c*) Information response rate will be a "stronger" function of hours of deprivation (i.e., the slope will be greater) than will complexity response rate.

The hypotheses concerned with incentive interpretations were: (*a*) Response frequency will be an increasing monotonic function of the information value of light series. (*b*) Response frequency will be an increasing monotonic function of the complexity value of light series. (*c*) Response frequency will be a stronger function of information value than of complexity value.

Method

Thirteen undergraduate male *S*s, each paid $10.00, were visually deprived for 8 hr. Each *S* was provided with a control box containing a button and a five-position dial. The button activated electronic programing apparatus which caused series of faint red and green light flashes to appear in the ceiling. Each series consisted of six flashes at 1-sec. intervals. The dial positions corresponded to differing values of information and complexity. Three relative values of each dimension were employed, .00, .50, and 1.00 (see Table 1), with 1.00 denoting the maximum value. As noted above, zero information and complexity are represented by identical series—repeated flashes of a single color. Consequently, only five series are required to provide the three values of both dimensions. The order with which the series were associated with dial positions was varied among *S*s, as was the color of the fixed flashes *a* or *b* (Table 1). So that *S*s might not forget the nature of the series associated with each dial position, a panel containing a tactile code for each series was attached to the base of the control box. As in the previous experiments cited, each *S* was given a 15-mg. Dexedrine spansule in an effort to prevent periods of sleep during the experimental session, since response functions over the 8-hr. period would have been difficult to interpret were uncontrolled fluctuations between sleep and wakefulness permitted.

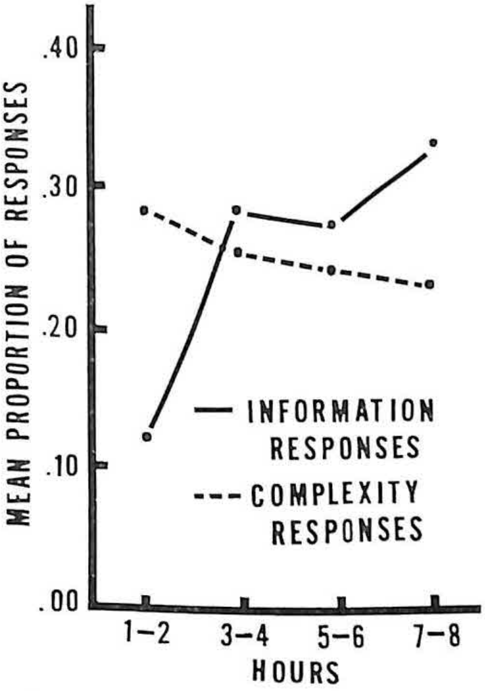


FIG. 1. Mean proportions of responses as a function of hours in the deprivation situation, excluding responses to series of zero value: Exp. I. ($N = 13$.)

The experiment was introduced as a study of boredom. A thorough explanation was made of the operation of the apparatus and of the nature of the light series associated with each dial position, which *S* was required to memorize. The *S*s were told that they might press the button whenever they wished but that there was no requirement to respond at all. They were instructed in the use of the tactile code and asked to refer to it whenever their memorization might falter. Further procedural details were the same as those reported in Exp. II of the prior study (Jones et al., 1961).

Results

In Fig. 1, response functions over hours of deprivation are plotted separately for information and complexity series. The data are the proportions of total responses tabulated over 2-hr. periods. Thus, for each *S* there are four values summing to 1 for each function. For this comparison, responses to the single series containing zero information and complexity were not included. Thus for each 2-hr. interval the proportion represents the pooled responses to the .50 and 1.00 information (or complexity) series. Data were analyzed by the analysis of variance of trends (Grant, 1956). The increasing linear trend of response to information series was significant, $F(1, 12) = 7.17$, $p < .01$. No significant trends were demonstrated for the complexity function. Further, the two functions were shown to differ significantly by a *t* test based upon the linear components of the data, $t(12) = 2.86$, $p < .02$. Figure 2 shows response functions plotted separately over the increasing values of the information and complexity series. Data are the proportions of total responses tabulated over the values .00, .50, and 1.00. Thus for each *S* there are three proportions summing to 1 for both functions. Responses to the zero condition are common to both functions but have different proportional representation when the functions are considered separately, as may be noted in Fig. 2. The analysis of variance shows a significant linear trend over information values, $F(1, 12) = 24.65$, $p < .001$, but a significant quadratic trend over complexity values, $F(1, 12) = 25.60$, $p < .001$. The difference between the information and complexity trends, assessed by a *t* test for correlated means based upon the linear component "scores" for *S*s, was found to be significant at the .01 level, $t(12) = 12.19$.

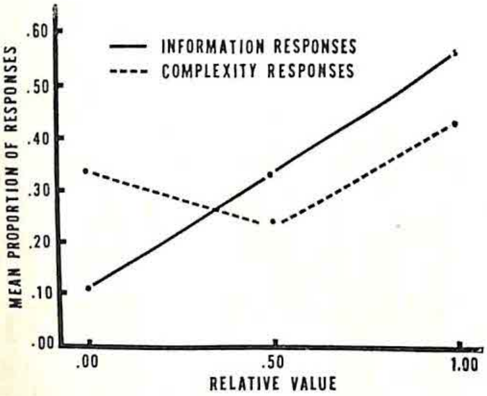


FIG. 2. Mean proportions of responses as a function of the relative value of the information and complexity series: Exp. I. ($N = 13$.)

EXPERIMENT II

In Exp. I, with information and complexity incentives equally available, response to the complexity series failed to show an increasing function over hours of deprivation or incentive values, although both such functions were demonstrated with respect to information series. It seemed possible that the information drive-incentive relationships might obscure complexity drive-incentive relationships when both are operating, but that the latter might manifest themselves when information is held at zero. Experiment II was designed to test this possibility. It was hypothesized that, in a condition of visual deprivation characterized by the absence of information incentives, the frequency of response for complexity series will be an increasing monotonic function of (*a*) the number of hours of deprivation, and (*b*) the complexity value of the series.

Method

College students served as paid *S*s. The method was identical to that of Exp. I, except for the light series; the five complexity values .00, .25, .50, .75, and 1.00 were employed (with zero information value in each case). Each series consisted of eight flashes, the smallest number providing for the consistent construction of all five complexity values.

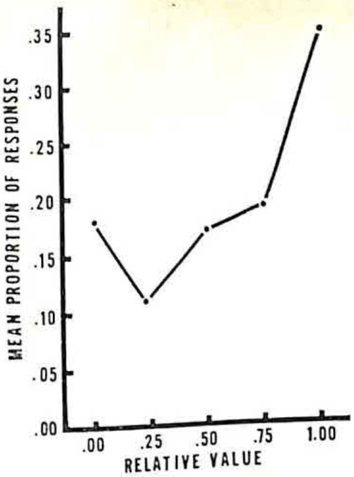


FIG. 3. Mean proportions of responses as a function of the relative complexity value of the stimulus series: Exp. II. ($N = 12$.)

Results

The data were first tabulated as the proportions of total responses which were made during successive 2-hr. periods of deprivation. An analysis of variance of trends performed on these data again showed no significant trend, linear $F\,(1, 11) = 1.40, p > .05$. Figure 3 shows the response function plotted over increasing values of complexity. As before, data are in the form of proportions of responses made to each value. Thus for each S there are five proportions summing to 1. In contrast to the findings of Exp. I, the analysis of variance showed both the linear and quadratic components of the function to be significant (linear $F = 4.91$, quadratic $F = 6.41$; $df = 1/11$ and $p < .05$ in each case).

EXPERIMENT III

The purpose of Exp. III was to assess the relative strengths of the drive-incentive relationships associated with stimulus complexity and stimulus fluctuation when both classes of incentives are equally available. In view of the results of Exp. II, it was expected that there would again be no evidence of an increasing trend of complexity responses over hours of deprivation, and that, since complexity is superordinate to fluctuation with respect to the statistical properties of sequences of stimuli, there would similarly be no evidence for an increasing trend of fluctuation response over hours of deprivation. Regarding incentive relationships, it was hypothesized (*a*) that response rate would again be an increasing linear function of complexity value of the series, (*b*) that response rate would be an increasing function of fluctuation value of the series, and (*c*) that the response function over complexity values would be steeper than that over fluctuation values.

Method

The method was identical to that of Exp. I except for the light series available to the 12 *Ss*. Complexity and fluctuation series were represented by the relative values .00, .50, and 1.00. Since the zero values of complexity and fluctuation are represented by identical stimulus series—repeated flashes of a single, fixed color—only five series are required to provide the three values of both dimensions. Each series consisted of eight flashes.

Results

In Fig. 4, response functions over hours of deprivation are plotted

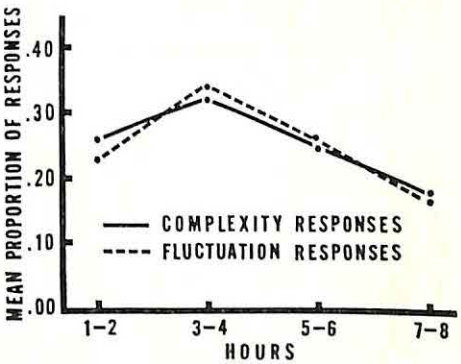


FIG. 4. Mean proportions of responses as a function of hours in the deprivation situation, excluding responses to series of zero value: Exp. III. ($N = 12$.)

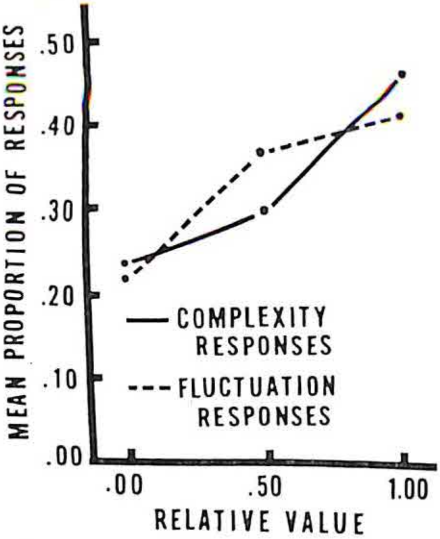


FIG. 5. Mean proportions of responses as a function of the relative value of the complexity and fluctuation series: Exp. III. ($N = 12$.)

separately for complexity and fluctuation series; data are the proportions of total responses tabulated over 2-hr. periods. Analysis of variance shows complexity responses to be a significant *decreasing* linear function of hours of deprivation, $F\ (1, 11) = 5.67$, $p < .05$, and thus necessarily significantly different from the increasing trend of information responses over hours found in Exp. I; the similar appearing downward trend of fluctuation responses, however, was not significant (linear $F = 1.94$, quadratic $F = 2.96$; $df = 1/11$ and $p > .05$ in both cases), although the function also differed significantly from the information response function over hours, $t\ (23) = 2.44$, $p < .025$. Figure 5 shows response as a function of complexity and fluctuation values. Data are the proportions of total responses tabulated over the values .00, .50, and 1.00, separately for each function. Analysis of variance showed both functions to be significantly linear at the .05 level—for complexity function, $F\ (1, 11) = 7.86$; for fluctuation function, $F\ (1, 11) = 8.03$. The difference between the slopes of the functions was not significant ($t < 1.0$).

EXPERIMENT IV

The purpose of the final experiment was to assess the drive-incentive relationships associated with stimulus fluctuation when the superordinate complexity dimension (and necessarily the information dimension) is held at zero. In Exp. III, neither complexity nor fluctuation responses showed the hypothesized trend over hours. In view of these results, and in view also of the previous negative findings with respect to a complexity response function over hours of deprivation when information was held at zero (Exp. II), it seemed unlikely that fluctuation responses would show an increasing trend over hours of deprivation when complexity is held at zero. It is theoretically possible that such a function might exist, however, despite the presumed transitivity of such effects, so that an empirical test was felt to be required.

In accord with the results of Exp. III, it was hypothesized that response rate would be an increasing linear function of the relative fluctuation value of the light series. It will be remembered, however, that in the case of complexity series, response functions over values had somewhat different properties depending on the presence (Exp. I) or absence (Exp. II) of a competing stimulus dimension. The present experiment provides for the demonstration of a similar phenomenon with respect to fluctuation, should such exist.

Method

The method was identical to that of the preceding experiments, except for the light series made available to the 11 *Ss*. The five fluctuation values .00, .33, .50, .67, and 1.00

were employed (complexity and information at zero in each case). Each series consisted of 12 flashes, the smallest number providing for the consistent construction of all five values. The .00, .50, and 1.00 values are given in Table 1. The .33 and .67 values refer to series in which one third and two thirds of the flashes are alternations of color with respect to the immediately preceding flash.

Results

The data were analyzed in the same manner as in the preceding experiments. The analysis of variance showed no significant trends of proportions of responses over hours of deprivation (linear $F < 1.0$). Figure 6 shows the response function plotted over increasing values of fluctuation. The analysis of variance of trends showed the quadratic component of the function to be significant; the linear component was not significant, quadratic $F\ (1, 10) = 6.75,\ p < .05$.

DISCUSSION

Drive hypotheses.—The results of the four experiments support the view that visual information deprivation functions as a drive variable, but fail to support drive interpretations of visual complexity and fluctuation deprivation, the term drive being used here in the sense of D in the Hull-Spence formulations (Hull, 1952; Spence, 1956). It had been hypothesized with respect to each of the three variables that response rate would be an increasing function of hours in the deprivation situation, where *S*s were permitted free response from the beginning. This hypothesis, in order to bear directly on a drive interpretation, requires the assumption that the absolute magnitude of the incentive is in each case very small, so that the progressively energizing effect of successive hours of deprivation will not be offset by satiation occurring during the experimental session. Positive results necessarily corroborate this assumption, while negative results are indeterminate. With regard to the information variable, this assumption was supported in the prior study (Jones et al., 1961) and again in Exp. I of the present paper by the significant linear trend of information-producing responses over time. Thus, the information incentive was of sufficiently low magnitude that *S*s showed no response decrement attributable to satiation. The hypothesis of increasing response rate over time was not supported with either the complexity or fluctuation variables, however. Two interpretations are possible: (*a*) that there is no drive variable associated with complexity or fluctuation deprivation, or (*b*) that there is a drive variable but that it is extremely weak relative to that associated with information deprivation, so that such minute incentives as were comprised by the complexity and fluctuation light series were sufficient to produce satiation. (It should be acknowledged that if complexity and fluctuation response rates had been found to increase over time, the demonstration of drive properties would have required a further experimental step to eliminate the possibility that the increased response rates represented increments in habit only, rather than increments in drive.)

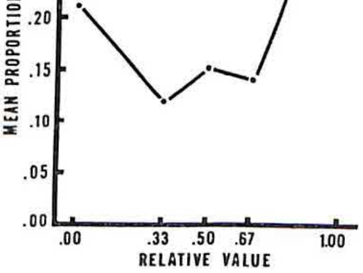


FIG. 6. Mean proportions of responses as a function of the relative fluctuation value of the stimulus series: Exp. IV. ($N = 11$.)

Incentive hypothesis.—In the prior study by Jones et al. (1961), response rate was shown to be an increasing monotonic function of the magnitude of the information incentive, a finding consistent with an incentive motivational interpretation as developed in the Hull-Spence view of reinforcement theory. This finding was replicated in Exp. I of the present study.

Since information series contain nonzero values of complexity as well, it appeared possible that the incentive properties of the series were a function not of the information associated with them, but of their physical randomness or complexity. Were this true, the slope of the response function over complexity values would equal that over information values. Experiment I was designed to clarify this issue. As is shown in Fig. 1, the functions are quite disparate; the information function is almost entirely linear, while the complexity function is neither linear nor monotonic. Thus an incentive-motivational interpretation of the information dimension seems clearly established as independent from whatever incentive properties may be associated with complexity.

The results of Exp. I demonstrate the dominance of the information incentive motivational principle over that associated with complexity when both classes of incentives are equally available to *S*s. Experiment II provides an assessment of the complexity incentive with information held at zero; a comparison of Fig. 1 and 3 shows that the shape of the function does in fact change in the direction of increased linearity. In Exp. I, only the quadratic component is significant. The function remains nonmonotonic, however, due to the relatively high frequency of response to .00 complexity; from .25 complexity upward the function is monotonic. Thus, the strength of the incentive-motivational properties of the complexity dimension appears to vary inversely with the presence of simultaneously available information incentives.

Complexity series necessarily contain fluctuation values, just as information series contain complexity values, and a parallel question concerning the independence of the complexity incentive dimension is raised. If the incentive value of complexity series were simply a function of their associated fluctuation rather than their physical randomness, response functions over equal values of complexity and fluctuation would have equal slopes, whereas the hypothesized dominance of the complexity incentive dimension would be shown by a steeper slope. The results of Exp. III (Fig. 5) showed no significant difference between the two slopes. Thus it appears that stimulus complexity does not constitute an independent incentive dimension; the incentive properties associated with physically random stimulus series are interpreted as due solely to the fluctuation of stimulus categories rather than to the attribute of randomness.

Experiment IV assesses the fluctuation function with complexity held at zero and with five rather than three values of series. Responses to this more finely graded set of incentives are distributed quadratically and nonmonotonically (Fig. 6) in a fashion very similar to the complexity function obtained in Exp. II plotted over five values. This correspondence would be expected in view of the interpretation above that the incentive properties of stimulus complexity are due solely to the associated fluctuation of stimulus categories.

In the discussion of Exp. I and II, it was noted that the incentive properties of the complexity dimension appear to vary inversely with the presence of simultaneously available information incentives. In view of the evidence indicating that the incentive properties of complexity series are in fact reducible to those of fluctuation, it would seem necessarily to follow that fluctuation series—in which complexity is held at zero—would show the same inverse relationship; i.e., the linear component of the fluctuation incentive function would be steeper in the absence of information series (Exp. IV) than when "competing" information incentives were simultaneously available. A simultaneous comparison of information and fluctuation variables was not included in the above experiments. As a control condition, however, Jones et al. included a 1.00 fluctuation series with the set of information series, all equally available to *S*. Response rate for the 1.00 fluctuation series did not differ significantly from that for the series which represented .00 value of both information and fluctuation, resulting in an almost flat fluctuation gradient plotted on those two points. Thus the expected transitivity of the incentive functions associated with the three dimensions appears to be verified.

REFERENCES

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

JONES, A. Supplementary report: Information deprivation and irrelevant drive as determiners of an instrumental response. *J. exp. Psychol.*, 1961, **62**, 310–311.

JONES, A., WILKINSON, H. J., & BRADEN, I. Information deprivation as a motivational variable. *J. exp. Psychol.*, 1961, **62**, 126–137.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received June 3, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 432–438

CUE TRIALS, FREQUENCY OF PRESENTATION, AND MEDIATING RESPONSES[1]

MARGARET JEAN PETERSON

Indiana University

The effects of variations in the strengths of acquisition-pair associations upon verbal-mediating responses and response availability were investigated by (a) inserting cue trials consisting of the stimulus alone into a mediation paradigm: A-B, A, B-C, B, A-CDE; and by (b) varying acquisition-stage frequency. Ss were to give the response on cue trials and to select the response which made the "best pair" with A on the test trials. Mediation was shown if test-trial selections of C following mediation presentation exceeded those following a control presentation which replaced B with a new stimulus in Stage 2. When cue-trial responses were correct, mediation was evidenced and did not increase as a function of the acquisition-stage frequency. Response availability was reduced to a near-chance level.

In the past, most *E*s dealing with verbal mediation have required that the two acquisition lists be learned to a higher level of proficiency than is customary in paired-associate learning experiments (e.g., Bugelski & Scharlock, 1952; Russell & Storm, 1955, used a criterion of three correct trials). Presumably, the assumption was that overlearning the acquisition lists would facilitate the establishment of mediating chains.

Using a different technique, Peterson and Blattner (1963) found that test-trial responses following mediation presentation increased at the same rate as test-trial responses following control presentation when the frequency of presentation of the acquisition stages was increased from 1 to 6. Similarly, Horton and Hartman (1963) reported little evidence of mediated facilitation as a function of what they called "exposure frequency" of the acquisition presentations.

The experiments reported herein were conducted to investigate the relationships between acquisition-pair strengths and verbal mediation. Strength of the acquisition-pair associations was manipulated by (*a*) presenting the acquisition stages one, three, or six times, and by (*b*) interpolating cue trials on which the stimulus alone was presented following display of the acquisition stimulus-response pair and *S* was to provide the appropriate response. If a verbal-mediating chain were more likely to be formed when the acquisition stages were well established, an increase in mediated facilitation would be expected as a function of increased frequency of acquisition-stage presentation and of correct cue-trial responding. Experiment I was conducted to test the predictions. It should be noted that mediated facilitation may reflect an increase in the number of test-trial responses following mediation presentation or a decrease in the corresponding number following control presentation or both. If responses on cue trials were incorrect, mediated facilitation should be reduced.

In a study mentioned previously

---

[1] This research was supported by Grant M5209 from the National Institutes of Mental Health. The assistance of Keith Blattner in the conduct of Exp. I and II is gratefully acknowledged.

TABLE 1

SAMPLES OF THE LEARNING MATERIALS

Frequency of Presentation	Stage		Type of Presentation	
			Mediation	Control
1	Acquisition			
	First	A-B	QJH XZF	QJH XZF
	Cue	A	QJH	QJH
	Second	B-C	XZF MWB	ZGW MWB
	Cue	B	XZF	ZGW
	Test Trial	A-CDE	QJH MWB FHJ GXM	QJH MWB FHJ GXM
3	Acquisition			
	First	A-B	QJH XZF	QJH XZF
		A-B	QJH XZF	QJH XZF
		A-B	QJH XZF	QJH XZF
	Cue	A	QJH	QJH
	Second	B-C	XZF MWB	ZGW MWB
		B-C	XZF MWB	ZGW MWB
		B-C	XZF MWB	ZGW MWB
	Cue	B	XZF	ZGW
	Test Trial	A-CDE	QJH MWB FHJ GXM	QJH MWB FHJ GXM
6	Acquisition			
	First	A-B	QJH XZF	QJH XZF
		A-B	(QJH XZF shown 5 more times)	(QJH XZF shown 5 more times)
		⋮		
	Cue	A	QJH	QJH
	Second	B-C	XZF MWB	ZGW MWB
		B-C	(XZF MWB shown 5 more times)	(ZGW MWB shown 5 more times)
		⋮		
	Cue	B	XZF	ZGW
	Test Trial	A-CDE	QJH MWB FHJ GXM	QJH MWB FHJ GXM

(Peterson & Blattner, 1963) some evidence suggested that mediational chains were established by a single presentation of each of the acquisition stages. Then increases in mediation found as a function of increased acquisition-presentation frequency could reflect incremental strengthening of the mediational chains, or the additional opportunities for the chain to be established in an all-or-none fashion. If acquisition stages were presented until cue trials always ended in a correct response, an increase in mediational facilitation might be expected if incremental strengthening occurred as the result of overlearning the acquisition pairs but would not necessarily be expected if the mediating chain were established in an all-or-none manner. In Exp. II, frequency of acquisition-stage presentation was varied as in Exp. I but if *S* erred on a cue trial the pair was reshown until he emitted the correct response.

Experiment III was essentially a control experiment to evaluate methodological differences between Exp. I and II.

GENERAL METHOD

Materials and design.—Consonant trigrams (CCCs) of from 0 to 17% association value (Underwood & Schulz, 1960) were used to construct mediation sets (Table 1). Using memory-drum (Exp. I) or card (Exp. II and III) presentation, a mediation set began with the first-stage presentation of a pair of consonant trigrams, A-B. Stimulus A was then shown alone as a cue trial and *S* was instructed to read off the stimulus and to give the response which had been paired with it on the previous presentation. In the second acquisition stage, the response term from the first pair, B, was shown as the stimulus with another trigram-response term, C, which was then followed by the appropriate cue trial, B. In the test stage, the first stimulus trigram, A, was presented with the second-stage response, C, and two additional trigram-response alternatives, D and E. For memory-drum presentation of Exp. I the four trigrams were typed in a row. The *S*s were told that the left-hand term was the stimulus to be paired with each of the other response terms. The stimulus was printed above the row of three response alternatives for card presentation of Exp. II and III. (Additional details specific to the manner of presentation used for each experiment are described with that experiment.) The *S*s were instructed to verbally pair each of the response trigrams successively with the stimulus trigram, A, and to then select the response alternative (i.e., C, D, or E) which made the "best pair or seemed right" with Stimulus A. A control set was prepared for each mediation set by substituting a new extraneous trigram for Stimulus B in the second-stage pair, X-C, providing a total pool of 24 mediation and their respective 24 control sets.

The first- and second-stage pairs were presented one, three, or six times before the cue trial was displayed. Thus the experimental design consisted of a 2 × 3 factorial in which type of presentation (mediation vs. control) was orthogonal to frequency of presentation (1, 3, or 6). The resulting six cells were each represented four times among the 24 sets assigned to a particular *S*. The order of presentation of the sets so assigned was randomized.

Procedure.—The *S*s were instructed:

> In this experiment you will be shown some consonant syllables. A consonant syllable is composed of three consonant letters. Soon you will see a pair of consonant syllables. You are to read the letters aloud. Then the first syllable will be shown again and you read off the letters and give the letters of the consonant syllable just paired with the first syllable. Soon you will see another kind of presentation. Four consonant syllables will be shown. The one on your left (on top) is the stimulus syllable and you are to read it off with each of the other three syllables, the response syllables. For example, if the stimulus were ABC and the responses were DFG, HJK, and LMN, you would say ABC DFG, ABC HJK, ABC LMN. You are to decide which of the last three response syllables makes the best pair or seems right with the stimulus. Do you have any questions? Let's try a few.

Prior to the experimental run a practice set was given at each of the three levels of frequency. The practice sets differed from both mediation and control sets in having different stimulus and response trigrams on the acquisition and test trials. The trigrams used in the practice sets were not part of the materials employed in the experimental pool described above. After the experimental session, *S*s were questioned about the method of selection of responses on the test trial and dismissed. None of the *S*s appeared to have insight into the experimental design.

Subjects.—A total of 54 introductory psychology students at Indiana University participated to fulfill course requirements, 18 in each experiment with no duplication of *S*s from experiment to experiment.

EXPERIMENT I

Method

The materials were typed in capital letters on a long memory-drum tape and were displayed at a 4-sec. rate. The drum was stopped on cue and on test trials until *S* responded or indicated that he was unable to respond (cue trials).

Results and Discussion

Although more C selections were made following mediation presentation than following control presentation, the difference was short of statistical significance, $F_{\text{Type of Presentation}}$ (1, 17) = 4.00, Table 2. The *F* for the main effect of frequency of presentation was less than unity. The third main effect, correctness of cue-trial responses, was reliable beyond the .001 level of significance, *F* (1, 17) = 45.46, showing that when cue-trial responses were correct, C was selected more frequently on the test trials than when the cue-trial responses had been incorrect.

Primary interest was in the interactions. As predicted, when cue-trial responses were correct, mediational facilitation was shown by the reliably greater number of Test-Trial C selec-

TABLE 2

Analyses of Variance, Number of Test-Trial C Selections

Source	df	Exp. I	Exp. II	Exp. III
Subjects (S)	17			
Type of presentation (T)	1	4.00	63.84***	16.10***
Frequency of presentation (F)	2	—	—	1.34
Cue-trial responses (C)	1	45.46***		68.70***
T × F	2	—	—	1.97
T × C	1	5.39*		5.65*
F × C	2	6.14**		9.89***
T × S	17		—	
F × S	34		—	
C × S	17			
T × F × C	2	—		—
T × F × S	34	—		—
T × C × S	17	—		2.47*
F × C × S	34	1.73		1.42
Error	34	(23.1573)	(30.1112)	(13.8240)
Total	215		(df = 107)	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

tions following mediation than control presentation ($t = 2.78, p < .02$, Table 3). The difference following incorrect cue trials was not reliable ($t = 1.25, p > .05$). Selection of the C alternative on test trials was more frequent when cue-trial responses were correct than when they had been incorrect for both mediation ($t = 13.44, p < .001$) and for control presentation ($t = 9.42, p < .001$). The above t tests were computed using the T × C × S interaction term with $df = 17$.

In an earlier study (Peterson & Blattner, 1963) the number of C alternatives selected on the test trials differed from an a priori chance level. This phenomenon, response availability, was assumed to reflect some kind of enhancement of C on the test trial deriving from its immediately precedent exposure during the second acquisition stage. In the present experiment, the number of C selections on the test trials following all control presentations exceeded chance expectations, $t\ (17) = 4.25$, $p < .01$. When control presentations were separated into those with correct cue-trial responses and those with incorrect, test-trial selections did not differ from chance in the first case, $t\ (17) = 2.05$, $p > .05$, but were reliably below chance for incorrect cue trials, $t\ (17) = 6.99$, $p < .01$. Correct cue trials appeared to facilitate the formation of mediating chains and to reduce response availability to a chance level.

The significant F × C interaction (Table 4) arose from a general increase over frequency of presentation when cue trials were correct, $F\ (2, 50) = 8.64$, $p < .001$, and a decrease as

TABLE 3

Mean Test-Trial C Selections as a Function of Type of Presentation and Cue-Trial Responses

	Cue-Trial Responses	
	Correct	Incorrect
Exp. I		
Mediation	6.06	1.22
Control	5.06	1.67
Exp. III		
Mediation	6.50	1.22
Control	5.06	1.72

intervening stimuli, a result which is strongly supported by the data of Hunt (1961).

These earlier studies of memory in concept learning are incomplete in at least one sense which is the primary issue explored in the present experiments. Simultaneous and successive stimulus-presentation methods are end points on a quantifiable dimension, to be described as stimulus availability, which may be operationally defined in terms of the number of previously presented stimuli to which S has access on any trial. The purpose of the present experiments is to evaluate the effect of systematic variation in the amount of previous information available to S. In light of earlier results, a monotonic improvement in performance over intermediate conditions of availability is expected.

EXPERIMENT I

Method

Procedure.—Stimuli used in this experiment were geometric designs printed on 3 × 3 in. stiff white cardboards. Each of 64 cards contained one design formed by the combinations of levels within the following six binary dimensions: color (red-green), size (large-small), texture (speckled-solid color), number (one-two figures), form (circle-ellipse), and border (connected-broken).

The procedure was patterned after that of Hovland and Weiss (1953) and Cahill and Hovland (1960). Preliminary instructions to S about the concept-learning task included (*a*) a description of all the stimulus dimensions and their levels, (*b*) an indication of the number of dimensions which might be relevant to problem solution, and (*c*) an explanation of the information conveyed by positive and negative instances of the concept. After instructions, *S*s were given a practice problem which involved learning a conjunctive concept with one level on each of two relevant dimensions, e.g., *speckled* figure within a *broken* border. In the experiment proper, each S was given a series of three conjunctive concept problems to solve. During any problem, the stimuli were placed before S one at a time, E describing each as a positive or negative instance of the concept. Within 15 sec. after the presentation of each stimulus, S was required to verbalize a guess about the concept, thus providing E with a trial-by-trial record of his hypotheses. The S was not permitted to see this record until problem solution had been attained. A problem was terminated after S stated the correct concept on four consecutive trials. Approximately 1 min. passed between successive problems during which S's questions were answered.

Within each problem, stimulus order was random with two restrictions: (*a*) the first card in every problem was a positive instance and (*b*) all possible positive and negative instances within a given problem were presented once before any was repeated.

Experimental design.—An independent group of 12 *S*s served under each of six conditions of availability of previously presented information. For each of these groups, a number of previously presented stimuli, viz., 0, 1, 2, 3, 4, or 5, remained before S as he responded to each new stimulus. In Cond. 5, for example, stimuli which had been presented on the five preceding trials (except for the first five trials of the problem), plus the particular instance exposed on that trial, were available to S for inspection. Each time a new stimulus was presented the earliest in the available series was removed. The E presented each card such that positive instances appeared in a row on the table closer to S than negative instances to permit immediate identification of the status and information conveyed by any particular stimulus. In Cond. 0, each stimulus card was removed before the next was presented.

Each of the three experimental problems given to S involved a different number of dimensions (one, two, or three) relevant to solution; the actual dimensions used as relevant varied between problems in the series. In all problems, three irrelevant dimensions were permitted to vary. Two sets of three problems were constructed, each of which was used for half the 12 *S*s in each availability condition; all six possible orders of three problems in a set were used, one order for each of six *S*s.

Subjects.—The *S*s were 72 volunteers from introductory psychology courses[2] and were assigned to groups randomly as they appeared. Each S served individually in a single session lasting no longer than 45 min.

---

[2] In Exp. I and II, *S*s were undergraduates at the University of California, Berkeley. The *S*s in Exp. III were undergraduates at the University of Utah.

TABLE 1

Analyses of Variance on Number of Instances to Solution: Exp. I

Source	*df*	Total Instances		Negative Instances		Positive Instances	
		MS	*F*	*MS*	*F*	*MS*	*F*
Availability (A)	5	892.76	15.25**	684.98	25.12**	14.51	6.85**
Linear	1	3,519.24	60.11**	2,715.66	99.57**	52.00	24.54**
Quadratic	1	722.38	12.34**	563.17	20.65**	9.90	4.67*
Cubic	1	70.94	1.21	61.45	2.25	0.34	0.16
Quartic	1	36.57	0.63	26.36	0.97	0.83	0.39
Quintic	1	114.68	1.96	58.26	2.14	9.46	4.47*
*S*s/A	66	58.55		27.27		2.12	
Relevant attributes (R)	2	9,823.91	176.89**	8,417.18	170.07**	54.34	27.88**
Linear	1	18,225.00	328.15**	15,604.17	315.29**	101.67	52.22**
Quadratic	1	1,422.82	25.62**	1,230.19	24.86**	7.00	3.59
A × R	10	625.50	11.26**	474.23	9.58**	11.15	5.72**
Lin. × Lin.	1	4,005.26	72.12**	3,118.76	63.02**	55.37	28.41**
Lin. × Quad.	1	948.13	17.07**	733.72	14.83**	13.72	7.04**
Quad. × Lin.	1	852.54	15.35**	608.96	12.30**	20.44	10.49**
Quad. × Quad.	1	214.50	3.86	148.60	3.00	6.03	3.10
Cub. × Lin.	1	26.60	0.48	13.78	0.28	2.09	1.07
Cub. × Quad.	1	7.37	0.13	8.51	0.17	0.04	0.02
Quar. × Lin.	1	34.84	0.63	20.72	0.42	1.82	0.94
Quar. × Quad.	1	0.37	0.01	0.34	0.01	0.00	0.00
Quin. × Lin.	1	107.68	1.94	54.48	1.10	8.98	4.61*
Quin. × Quad.	1	57.68	1.04	34.40	0.70	2.99	1.54
Problem set	1	3.17	0.06	14.52	0.30	0.17	0.09
Sequence	5	88.60	1.60	71.94	1.45	1.72	0.88
Ordinal position	2	92.98	1.67	73.01	1.48	1.69	0.87
Residual	124	55.54		49.49		1.95	

* $p < .05$.
** $p < .01$.

Results

Summarized in Table 1 are three analyses of variance performed on total number of instances, number of positive instances, and number of negative instances presented to *S* prior to the four-trial criterion run. Separate analyses were performed in the expectation that these measures would show different effects, especially for number of relevant attributes since, as this variable changes, so do the relative numbers of positive and negative instances within the total set. Because there is considerable evidence indicating that negative instances are difficult to use in concept learning, it is possible that increasing the number of relevant attributes could produce a significant effect on performance measured by number of negative instances to solution without any attendant change in task difficulty or in performance measured in terms of number of positive instances to solution. Despite the fact that the number of negative and positive instances differed across problems, there is marked consistency in outcome among the three analyses.

The main effects of availability and number of relevant attributes and the interaction of these variables were highly significant in all analyses. The nature of these effects is illustrated in Fig. 1. A significant quadratic component of the function for number of relevant attributes is consistent with earlier results (Walker & Bourne, 1961), showing the effect of increasing the number of nonredundant relevant dimensions to be a disproportionate increase in the difficulty of the task.

Availability of previous instances also affected performance nonlinearly; its facilitative effect decreased as its value increased. Indeed, at least in the case of problems with three relevant attributes, there may have been a decrement in performance with stimulus availability greater than four.

The reliable interaction of these variables reflects the fact that availability was less facilitative in problems of lesser complexity. In fact, there was little observable effect in problems with one relevant attribute in which solution was attained on the average in less than five trials. The significant components of interaction are consistent with the nonlinear main effects.

The analysis of positive instances differs from the others in two important ways: (*a*) only the linear component of the relevant-attributes variable was statistically significant and (*b*) significant quintic components occurred in the availability term and in the interaction of availability and relevant attributes. Neither outcome has an obvious explanation, although the latter finding is not highly significant and appears to be attributable to the irregularity of the performance-availability function in problems with three relevant dimensions in which positive instances are relatively few and their position in the stimulus sequence is particularly critical.

No other sources of variance were statistically significant, including the ordinal position effect which reflects a lack of general improvement in performance across three experimental problems.

EXPERIMENT II

The investigation of stimulus availability in Exp. I was limited in part

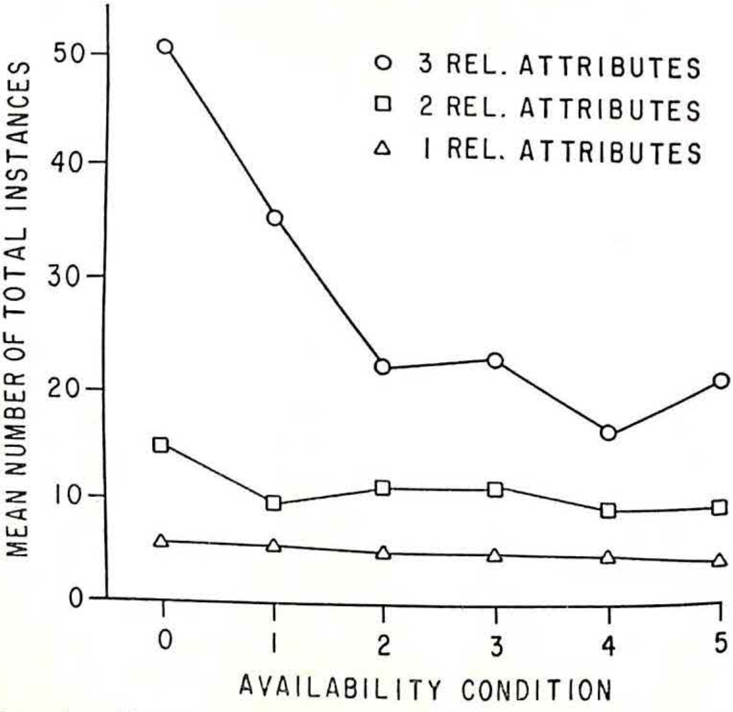


FIG. 1. Total number of instances to solution as a joint function of availability and number of relevant attributes. (Each plotted point represents the data of 12 *S*s: Exp. I.)

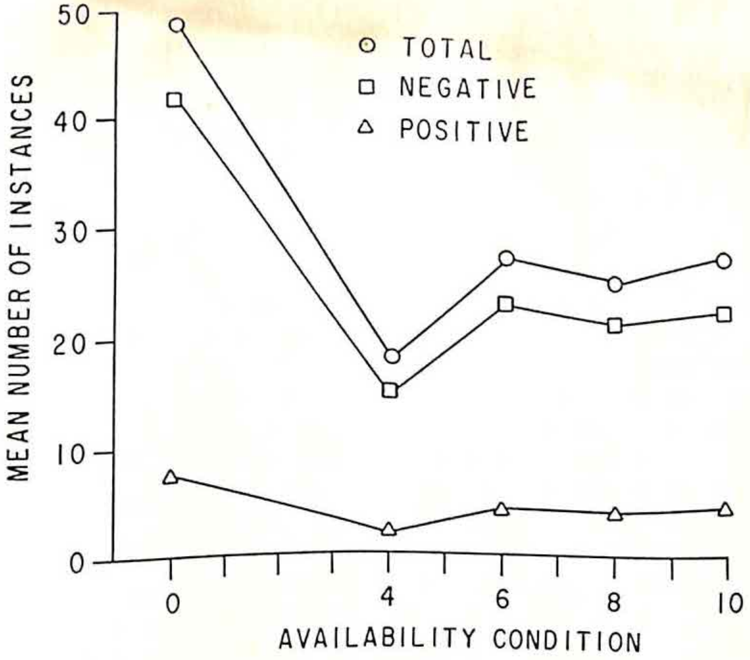


FIG. 2. Number of positive, number of negative, and total number of instances to solution as a function of availability. (Each plotted point represents the data of 12 *S*s: Exp. II.)

by the desirability of including parallel conditions at three different levels of task complexity and in part by lack of preliminary information about the significant range of this variable. A wider range of values was utilized in Exp. II in order to provide evidence on the reliability of the availability effect and to check on possible inversions in the function at higher values.

Method

In all respects except the following, the procedure was the same as that used in Exp. I. The *S*s were 60 volunteers from courses in introductory psychology,[2] assigned randomly but in equal number to five independent conditions of availability. These conditions were such that 0, 4, 6, 8, or 10 stimuli from trials immediately preceding the present one remained in view for *S*'s inspection. Because the effect of availability was largest in concept problems of greatest complexity in Exp. I, *S* was required to solve only one conjunctive concept involving three relevant attributes (*solid*, *red*, *ellipse*) and three irrelevant dimensions (size, number, and border).

Results

Trials to solution.—Plotted in Fig. 2 are the mean numbers of positive, negative, and total instances required by *S*s in the various availability conditions. Analyses of variance show availability to exert a significant effect on each measure (Total: $F = 11.46$; Positive: $F = 6.36$; Negative: $F = 11.40$; $df = 4/55$, $p < .01$). In order to assess the form of these performance-availability functions, scores for Availability Cond. 2 were estimated by linear interpolation between Cond. 0 and 4. Such estimates make the simplest possible assumption about the effect of availability and yield essentially the same values as those observed for Cond. 2 in Exp. I. In analyses on each of the three measures, only two components of trend, the linear and the quadratic, were significant (Total: $F_{\text{lin.}} = 19.79$, $F_{\text{quad.}} = 18.17$; Positive: $F_{\text{lin.}} = 12.04$,

$F_{\text{quad.}} = 9.29$; Negative: $F_{\text{lin.}} = 19.86$, $F_{\text{quad.}} = 18.59$; $df = 1/55$, $p < .01$). The outcome indicates that performance, no matter how measured, improves with stimulus availability, but that an optimal degree of availability exists at a small number of stimuli. Further analyses with Duncan's test support this conclusion, suggesting, as does Fig. 2, that the optimum is at or about four stimuli (in addition to the immediate stimulus of any trial). In all cases, Cond. 0 differed from Cond. 6, 8, and 10, which in turn differed as a group from Cond. 4 ($p < .05$).

Perceptual-inference and memory errors.—Hypotheses stated by *S* may be incompatible with information (*a*) in the immediate stimulus or (*b*) in the stimulus presented on any preceding trial. Cahill and Hovland (1960) termed erroneous hypotheses of the first type perceptual-inference (P-I) errors for they reflect either a failure to perceive a stimulus correctly or to draw proper inferences from it, and those of the second type memory (M) errors for they indicate a failure to recall and utilize correctly information in earlier stimuli. The analysis of erroneous hypotheses described by Cahill and Hovland was limited to a comparison of the two conditions of availability prescribed by successive and simultaneous stimulus presentation; thus it seemed appropriate to make a similar analysis on the more detailed evidence of the present experiment.

Table 2 shows the mean number of hypotheses in each availability condition which were incompatible with information in the stimulus presented on any trial, x, with information presented on the just preceding trial, $x - 1$, and so on back for eight degrees of remoteness from the immediately presented stimulus. All hypotheses which were incompatible with stimuli still exposed, e.g., x through $x - 4$ in Cond. 4, are labeled P-I errors, and those incompatible with previously exposed but no longer available stimuli, e.g., $x - 5$ through $x - 8$ in Cond. 4, are labeled M errors. Length of the trial series analyzed was limited in order to include at least one such series for all *S*s.

There were relatively few P-I errors; counting only those which were made to the immediately presented stimulus (Trial x), there was no reliable difference among the five availability conditions, $F(4, 55) = 1.74$, $p > .10$. The *S*s who made P-I errors on this trial, however, did have

TABLE 2

MEAN NUMBER OF ERRORS: HYPOTHESES INCOMPATIBLE WITH INSTANCES PRESENTED VARYING NUMBERS OF TRIALS EARLIER: EXP. II

Availability Cond.	Type of Error	Degree of Remoteness 0	1	2	3	4	5	6	7	8
0	P-I	1.00	—	—	—	—	—	—	—	—
	M	—	1.00	1.67	2.25	2.33	2.67	2.58	2.75	3.00
4	P-I	.25	.33	.33	.58	.42	—	—	—	—
	M	—	—	—	—	—	.58	.75	.75	.83
6	P-I	.68	.72	.68	.80	.80	.63	.75	—	—
	M	—	—	—	—	—	—	—	.58	.75
8	P-I	.92	.75	.75	.92	.75	.68	.70	.52	.53
	M	—	—	—	—	—	—	—	—	—
10	P-I	.58	.75	.67	.82	.87	.83	.88	.80	.85
	M	—	—	—	—	—	—	—	—	—

greater difficulty in concept learning. Subgroups were formed by dividing *S*s at the median number of P-I errors within each availability condition; these subgroups differed significantly in total trials to solution, $U(n_1 = 20,\ n_2 = 40) = 55,\ p < .01$. The result is consistent with findings of Cahill and Hovland; one factor important in producing overall performance differences appears to be the ability to utilize information from successive stimuli as they are presented in the trial sequence.

The *S*s in Cond. 4, 6, 8, and 10 had varying numbers of opportunities to make P-I errors with respect to stimuli presented earlier in the trial sequence. Since these groups differed in overall performance, an analysis of variance was performed on the number of hypotheses which were incompatible with information presented in remote but still exposed stimuli. These groups had four such trials in common, $x - 1$ through $x - 4$, and the analysis was limited to this series. Only the difference among groups reached an acceptable level of significance, $F\ (3, 44) = 3.48, p < .05$, with Cond. 4 producing fewer errors ($p < .05$) than any of the remaining groups according to Duncan's test. There was no significant trend across trials; hypotheses were equally liable to be incompatible with information in any still exposed stimulus, a result which is consistent with the performance of *S*s under simultaneous stimulus presentation (Cahill & Hovland, 1960).

Memory errors were numerous, particularly in Cond. 0 wherein *S* had to rely most heavily on retained information to delineate the correct concept. An analysis of variance was performed on the number of hypotheses which were incompatible with information in stimuli varying numbers of trials ($x - 1$ through $x - 8$) earlier in the sequence. This analysis was limited to a comparison of Cond. 0, whose errors over this sequence were all of the M type, and Cond. 8 and 10, whose errors were entirely of the P-I variety. Significant effects due to Conditions, $F\ (2, 33) = 4.15,\ p < .05$, and to the Conditions × Trials interaction, $F\ (14, 231) = 3.31,\ p < .01$, were observed, the latter confirming the apparent difference in trends over degrees of remoteness shown in Table 2.

Note that the mean numbers of M errors produced in Cond. 4 and 6 are smaller than that in Cond. 0. Such an outcome is not entirely unexpected in view of the fact that *S*s in these conditions had five or seven stimuli (a combination of positive and negative instances) for inspection on each trial. The formation and statement of a hypothesis compatible with available stimuli reduced the likelihood that it would be incompatible with stimuli no longer exposed.

EXPERIMENT III

Both Exp. II and the study by Cahill and Hovland (1960) demonstrate that only a small proportion of errors in concept learning is attributable to *S*'s failure to draw appropriate inferences from perceptually available information. More errors result from *S*'s apparent failure to remember earlier stimuli in such a way as to utilize their implications. That P-I errors may not be wholly unimportant, however, is indicated by a worsening of performance in complex tasks when more than five stimuli are available. A larger array of stimuli may yield some weak but significant interference effect with the implication that *S*s are confused by an excess of available information, fail to take it

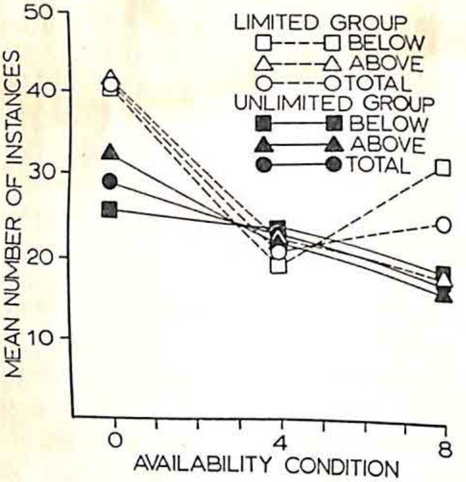


FIG. 3. Total number of instances to solution required (*a*) by the group as a whole and (*b*) by the two subgroups using more or less than the median time per trial to respond under six conditions of availability and response interval.

all into account, and as a result produce more errors or require more trials to solution. Alternatively, the outcome of Exp. II may be explained on the argument that *S*s were not given sufficient time to inspect and utilize all available information. A fixed, 15-sec. response interval was strictly enforced in all conditions of the preceding studies. The purpose of Exp. III was to determine the adequacy of these interpretations by testing the hypothesis that removal of the response-time limit will facilitate the utilization of available information and thereby reduce trials to solution under conditions in which more than four previous stimuli are available to *S*.

Method

With the following exceptions, the method was identical to that described for Exp. I. The *S*s were 120 volunteers from courses in introductory psychology,[2] assigned randomly and in equal numbers to 12 independent groups. Experimental treatments were formed by the factorial combination of three availability conditions (zero, four, and eight stimuli from immediately preceding trials remain exposed), two types of response interval (Limited to 15 sec. and Unlimited), and two different problems. Preliminary instructions were the same for all groups and did not indicate to *S*s that there was a limit on the permissible response interval. If on any trial *S* in the Limited condition had not stated a hypothesis within 15 sec., however, he was required to do so at that point. The length of the response interval for all *S*s on every trial was recorded. Each *S* solved one conjunctive problem with three relevant and three irrelevant dimensions; each of two problems, which differed with respect to the particular dimensions selected as relevant and irrelevant, was used for half the *S*s.

Results

Trials to solution.—Mean numbers of total instances to solution for the major conditions of the experiment are shown in Fig. 3. Analyses of variance performed on the three measures, number of positive, number of negative, and total number of instances to solution, produced nearly identical results. In all analyses, the linear component of availability was significant ($F_{\text{total}} = 14.03$; $F_{\text{positive}} = 14.68$; $F_{\text{negative}} = 13.59$; $df = 1/108$, $p < .01$). Variance due to response-interval conditions was large, but significant only with number of positive instances to solution as the performance measure, $F\,(1, 108) = 5.99$, $p < .05$. To test the hypothesis that a limited response interval will interfere with effective utilization of information in a large array of stimulus instances, *t* tests of the performance differences between Limited and Unlimited groups under Availability Cond. 8 were computed. All *t*'s were marginally significant ($t_{\text{total}} = 1.74$; $t_{\text{positive}} = 1.72$; $t_{\text{negative}} = 1.73$; $df = 38$, $p < .05$).

Response times.—Within each of the six major conditions of response interval and availability, *S*s were divided at the median into equal-sized subgroups composed of those who took more and those who took less time to respond averaged over all trials. Mean numbers of instances required by *S*s in these subgroups are plotted in Fig. 3. The largest difference between subgroups within a single condition appears with a Limited response interval at Availability Cond. 8. The difference between these subgroups is marginally reliable, $U(n_1 = n_2 = 10) = 29.5$, $p = .06$. No other subgroup comparisons were significant.

DISCUSSION

Data from the present study confirm and extend the conclusions of Hovland and Weiss (1953), Cahill and Hovland (1960), and Hunt (1961). They show that, in general, concept learning is significantly facilitated by the availability of previous stimuli, the degree of effect increasing at least up to some optimum. However, the availability dimension, represented in previous research only by the extreme conditions produced by simultaneous and successive stimulus presentation, is complexly related to performance. In the first place, availability had less effect in simpler problems, i.e., those with only one or two relevant attributes, which is probably attributable to the fact that such problems were solved rapidly. During the first few trials of any problem, there was little advantage to high availability because (*a*) full availability was present only after a prescribed number of trials (all groups started with one stimulus on the first trial) and (*b*) *S*s make better use of information in the initial stimuli, even though they may not remain available for inspection, than that in stimuli presented later in the series (Cahill & Hovland, 1960). Thus, the experimental procedure and a primacy effect both operate to facilitate solution to problems of lesser complexity, tend to obfuscate the effect availability has on M errors, and are sufficient to explain the significant interaction between task complexity and availability in Exp. I.

Secondly, the availability of more than an optimal number of stimuli from preceding trials may degrade performance, as shown in Exp. II. Results of Exp. III demonstrate, however, that this effect is probably attributable to limitations imposed on the length of the permissible response interval rather than to a perceptual-interference phenomenon brought on by an excess of stimulus information. When *S* is allowed unlimited time to respond to each stimulus, performance improves monotonically over availability conditions, at least to the point at which eight stimuli from previous trials are exposed for inspection. Moreover, as is evident in Fig. 3, *S*s limited to a 15-sec. response interval who took more than the median time to respond on each trial displayed at Cond. 8 performance which was numerically superior to that at Cond. 4. Thus, it appears that a certain critical amount of time may be necessary for inspecting and drawing appropriate inferences from each exposed stimulus and that, as the number of available stimuli increases, length of response interval must increase for maximally efficient performance.

The major role of availability, as expected, was in the reduction of M errors. Although the number of P-I errors did increase in availability conditions beyond the optimum, they were reliably less frequent than the number of M errors in Cond. 0. Hypotheses stated and assumedly tested by *S*s in this group clearly indicate that the majority of errors were attributable to a failure to recall previously displayed information. The fact that M errors (Cond. 0) but not P-I errors (e.g., Cond. 10) increase with increasing numbers of intervening stimuli, i.e., degree of remoteness, suggests an underlying short-term interference effect similar to retroactive inhibition (Hunt, 1961). Although plausible, confirmation of such interpretation will require other investigations in which, for example, variables such as stimulus similarity are systematically manipulated.

The effect of number of relevant attributes in Exp. I is consistent with the earlier findings of Walker and Bourne (1961). The procedure of the earlier study required *S* to categorize stimuli into two, four, or eight groups representing the number of independent conjunctive combinations of one, two, or three relevant binary dimensions. As such, problems were more difficult than in the present experiment where *S* needed only to respond to stimuli as positive or negative instances with respect to one combination of levels (the concept) on one, two, or three relevant dimensions. The fact that in both studies performance was a nonlinear function of number of relevant dimensions attests to the stability and generality of this relationship.

REFERENCES

CAHILL, H. E., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *J. exp. Psychol.*, 1960, **59**, 137–144.

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.

HUNT, E. B. Memory effects in concept learning. *J. exp. Psychol.*, 1961, **62**, 598–604.

WALKER, C. M., & BOURNE, L. E., JR. Concept identification as a function of amounts of relevant and irrelevant information. *Amer. J. Psychol.*, 1961, **74**, 410–417.

(Received June 10, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 449–457

SERIAL POSITION AND SEQUENTIAL ASSOCIATIONS IN SERIAL LEARNING[1]

WILLIAM F. BATTIG,[2] SAM C. BROWN,[3] AND MARY E. SCHILD
University of Virginia

12-item serial lists were compared, with 3 items in either same-adjacent, changed-adjacent, or same-nonadjacent serial positions as in a previous list, wherein the 3 List 2 cluster items appeared in either beginning (2-3-4) or middle (6-7-8) serial positions and were learned to either 1 or 3 correct cluster anticipations (1R-3R). Nonadjacent performance was significantly inferior to both same and changed conditions, which differed insignificantly. List 2 performance was better on List 1 middle than beginning cluster items, but 3R exceeded 1R only in Middle Position 8, indicating complex multiple-item associative units to develop gradually in middle positions, whereas initial items are learned by simple-position and sequential-adjacent associations.

Serial lists have long been assumed to be learned through the formation of direct and specific associations of each item with immediately adjacent (preceding and following) items, so that each item (except the first and last) comes to represent both a response to the preceding item and a stimulus for the following item (e.g., McGeoch & Irion, 1951, p. 90). Such a "chaining" conception has recently been seriously questioned due to the lack of positive transfer from prior serial learning to paired-associate learning of pairs of items which had appeared in adjacent positions in the serial list (Young, 1959, 1961, 1962), along with the findings that (*a*) reversal of item order in a serial list produces positive transfer only for middle items, where serial positions are most similar in the original and reversed lists (Young, Patterson, & Benson, 1963), and (*b*) performance is better on individual second-list items appearing in the same than in different serial positions from those of a previous list (Ebenholtz, 1963b; Young, 1962). These results have been interpreted as more consistent with a "position" hypothesis, whereby items are associated directly with their serial positions in the list rather than with each other. Particularly in view of indications that apparent differences in shape of the classic bowed serial-position curve were an artifact of the method of plotting these curves, and that this curve is invariant in shape over a variety of task and procedural conditions (McCrary & Hunter, 1953), it has been further suggested that this serial-position effect reflects the differential discriminability of various serial positions (Murdock, 1960; Schulz, 1955), and is thus entirely consistent with a position hypothesis of serial learning. Less direct evidence which is also more consistent with a position than a chaining explanation has come from studies of tasks involving significant modifications in the classical serial-learning procedure, through the use of spatial-position cues (Asch, Hay, & Diamond, 1960; Ebenholtz, 1963a; Voss, 1956) or of paired-associate tasks with a constant serial order

---

[1] This investigation was supported by Public Health Service Research Grant M-5769 from the National Institute of Mental Health to the senior author.

[2] Now at the University of Maryland.

[3] Now at Kansas State University.

(Battig, Brown, & Nelson, 1963; Brown & Battig, 1962).

However, considerable evidence contrary to serial-position associations as the principal mechanism of serial learning has recently come from (*a*) studies showing near perfect positive transfer when sequential order is maintained but absolute serial positions are modified (Rehula, 1960; cited in Underwood, 1963); and (*b*) demonstrations of significant positive transfer to paired-associate learning of pairs of adjacent items from a previous serial list, under conditions of strong associative connections between adjacent items or of minimal intralist associative strength (Horowitz & Izawa, 1963), and also with self-paced presentation conditions and explicit instructions concerning the nature of the relationships between the two lists (Jensen, 1962b). As a consequence, dual-process accounts of serial learning have also been suggested, consisting of associations involving both adjacent items and serial positions (e.g., Ebenholtz, 1963b; Underwood, 1963; Young, 1962). These processes have also been suggested to be of differential relative importance at the ends and in the middle of the list, although there has been disagreement over whether position associations are relatively more involved at the ends (Ebenholtz, 1963a) or the middle (Young, 1962). Still others have invoked additional processes representing associations of each item with a cluster of several preceding items (Horowitz & Izawa, 1963), the development of integrated multiple-item units proceeding from the beginning of the list as an "anchor point" (Jensen, 1962a), or the use of general overall information-processing strategies to determine the order in which items are learned (Feigenbaum & Simon, 1962).

The present study attempted an evaluation of the relative contribution to serial learning of some of the aforementioned processes, through the use of a transfer design wherein the second of two serial lists learned by each *S* included three items from the first list appearing under various conditions in (*a*) same-adjacent, (*b*) changed-adjacent, or (*c*) same-nonadjacent serial positions. To determine possible differential involvement of these various processes in different parts of the list and at different stages of learning, both level of learning and position (beginning or middle) of the critical second-list items within the first list were also varied systematically. Groups with clusters of three adjacent first-list items appearing in identical second lists in the same or changed serial positions could thereby be compared with each other, as well as with groups with isolated individual first-list items appearing in the same serial positions as in the first list, in order to evaluate the relative importance in serial learning of sequential associations between adjacent items, associations with serial positions, and of more complex higher-order learning processes.

METHOD

Materials.—Six 12-item serial lists, presented in Table 1, were constructed from 21 bigrams of minimal frequency of usage and associative strength (Underwood & Schulz, 1960). Two were learned as first lists (1A and 1B), both consisting of the same 12 items in different serial orders. The remaining 9 items, along with 3 of the List 1 items, constituted each of the four lists used as second lists. Interitem similarity was minimized by using no letter more than once as the first and second letter, while associative strength between the second letter of each item and the first letter of the following item was kept uniformly low and approximately equal across lists.

Design.—As shown in Table 2, 12 groups of 8 *S*s each were differentiated in terms of the particular set of lists learned as the first and second lists, and criterion of List 1 learning. Each of the two first lists was learned to a criterion of either one or three consecutive correct anticipations of the entire cluster of three items (CN-PD-WG) appearing in Adjacent Serial Positions 2-3-4 in List 1A and 6-7-8 in List 1B. The resulting four conditions on List 1 were each further subdivided into three conditions differing as to relative serial positions of the three common items in the two lists. Under the *same* condition, Cluster Items CN-PD-WG appeared in the same serial positions in List 2 as in List 1. Under the *changed* condition, these cluster items appeared in different adjacent positions in the two lists (e.g., items in Positions 2-3-4 in List 1A were shifted to Positions 6-7-8, and Items 6-7-8 in List 1B were shifted to Positions 2-3-4, by pairing List 1A with 2B, and 1B with 2A). The third *nonadjacent* condition provided that List 1 Items 3, 7, and 10 appeared also in these same positions in List 2, which was accomplished by pairing List 1A with 2C, and 1B with 2D. The complete design thus becomes a complete 2 × 2 × 3 factorial, with two first lists (1A and 1B), two List 1 learning criteria (one or three consecutive correct cluster anticipations, designated as 1R and 3R), and three relationships (same, changed, and nonadjacent) between serial positions of the three items common to both lists.

Subjects.—Each of 96 paid students or affiliates of the University of Virginia served for a single session of 1 hr. or less. Most *S*s had previous experience in one or more verbal-learning experiments, but none had used materials or procedures similar to those of the present study. A previously determined unsystematic sequence based on order of appearance in the laboratory was used to assign *S*s to the two first lists. However, within each List 1 condition, assignment of *S*s to 1R-3R or same-changed-nonadjacent groups was determined by number of trials to one correct cluster anticipation, so as to equate these in initial learning ability. Fulfilment of these matching procedures required the elimination of two additional *S*s, whereas three *S*s were discarded for failure to follow instructions or refusal or inability to learn List 1 to criterion.

TABLE 1

SERIAL LISTS USED FOR VARIOUS GROUPS

Position	First Lists		Second Lists			
	1A	1B	2A	2B	2C	2D
1	LQ	DZ	RH	VL	RH	VL
2	**CN**	NP	**CN**	BT	JS	BT
3	**PD**	**KX**	**PD**	ZF	**PD**	**KX**
4	**WG**	HC	**WG**	XM	VL	ZF
5	EJ	LQ	JS	RH	BT	XM
6	DZ	**CN**	VL	**CN**	ZF	RH
7	**KX**	**PD**	BT	**PD**	**KX**	**PD**
8	NP	**WG**	ZF	**WG**	XM	JS
9	HC	EJ	XM	JS	QE	QE
10	**GV**	**GV**	QE	QE	**GV**	**GV**
11	SR	SR	YB	YB	YB	YB
12	IW	IW	FK	FK	FK	FK

Note.—Items common to Lists 1-2 in bold-face type.

TABLE 2

LISTS LEARNED BY EACH GROUP

List 1 Cond.	List 2 Cond.		
	Same	Changed	Nonadjacent
Beginning			
1R	1A-2A	1A-2B	1A-2C
3R	1A-2A	1A-2B	1A-2C
Middle			
1R	1B-2B	1B-2A	1B-2D
3R	1B-2B	1B-2A	1B-2D

Procedure.—All lists were presented by standard memory-drum anticipation procedure, with a 2-sec. presentation rate and a 4-sec. intertrial interval. All *S*s were required to spell their responses. During a 1-min. break following achievement of the List 1 learning criterion, *S*s were told that a new list would be presented which might include some items from List 1. Learning of List 2 was then carried out to a criterion of two successive errorless trials.[4]

RESULTS

List 1 performance.—Analyses of variance revealed, as expected, that significantly more total correct responses, trials to criterion, and maximum correct responses on a single trial were required to reach the List 1 criterion by 3R than 1R groups, and by List 1 middle than beginning groups (all p's $< .025$). However,

---

[4] Three *S*s (each in a different group) failed to reach criterion on List 2. For these *S*s, performance measures were based on 71 trials, the minimum completed by all three prior to termination of the session.

List 2 same-different-nonadjacent groups did not differ significantly in any List 1 performance measure (all p's $> .05$). Significantly more correct responses on the three critical cluster items were made by the combined 3R than 1R groups ($p < .001$), but neither the difference between List 1 beginning-middle conditions nor List 2 same-different-nonadjacent conditions approached significance for this measure (all p's $> .10$). It can thus be concluded that the groups were both appropriately differentiated and matched with respect to subsequent List 2 comparisons, and that List 1 cluster items appearing also in List 2 had not been learned to different degrees by List 1 beginning and middle groups.

List 2 total performance.—Significantly more mean total errors on List 2 were made under the nonadjacent conditions (162.3) than under same (99.1) or changed (118.7) conditions, $F\ (1, 84) = 8.42$, $p \approx .005$, but the same and changed conditions did not differ significantly ($F < 1$). Although neither List 1 criterion (1R vs. 3R) nor List 1 cluster position (beginning vs. middle) produced significant overall effects on List 2 total errors, both the 1R-beginning (169.2) and 3R-middle (132.3) groups made more errors than the 3R-beginning (105.0) and 1R-middle (100.3) groups, leading to a significant interaction between List 1 criterion and cluster position, $F\ (1, 84) = 7.67$, $p < .01$. Since the 3R-middle groups had the most and the 1R-beginning groups the least practice on List 1, this finding indicates a nonmonotonic relationship of List 1 practice to overall List 2 performance, such that intermediate amounts of List 1 practice led to optimal List 2 performance. Neither List 2 cluster position nor any interaction approached significance in this analysis (all Fs < 1). Further error analysis revealed most of these (83%) to be omission errors for all groups, and showed no systematic or significant differences between groups in number or proportion of intrusion errors from either List 1 or 2, or in overt errors from neither list.

List 2 cluster performance.—Of primary interest in this experiment were comparisons between groups in relative performance in those cluster positions containing items which had also appeared in List 1. Due to marked individual differences within groups, along with the previously described significant group differences in overall List 2 performance, these comparisons were based on the percentage of total List 2 errors made on items appearing in the cluster positions. This percent error measure was calculated separately for each cluster position for each S in all groups. Also determined for all Ss in the same and changed groups were two additional measures more directly reflecting initial List 2 performance, designated as *before errors* (number of errors or trials prior to the first correct response to each cluster item), and *rank scores* (number of other List 2 items responded to correctly on one or more trials prior to the first correct response to each cluster item).

Presented in Fig. 1 are mean percent errors for each of the three cluster positions for both 1R and 3R criterion groups within both the same and changed conditions. Figure 1A presents results for Beginning Positions 2-3-4, based on List 2A, whereas measures for Middle Positions 6-7-8 (List 2B) appear in Fig. 1B. Also included in each figure for comparison purposes are the percent error measures from the appropriate combined 1R-3R nonadjacent conditions, based, respectively, on Beginning Positions 2-3-4

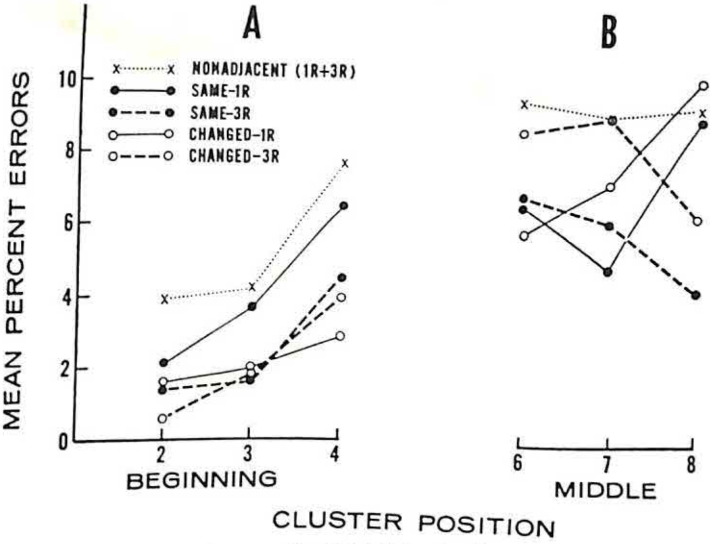


FIG. 1. Mean percent errors for List 2 beginning (A) and middle (B) cluster positions for each group.

from List 1C (Fig. 1A) and Middle Positions 6-7-8 from List 1D (Fig. 1B).

As shown in Fig. 1A, both same and changed conditions produced consistently fewer percent errors in the beginning cluster positions than did the nonadjacent condition. Analysis of variance revealed the superiority over the nonadjacent condition to be significant both for same and changed conditions, *F*s (1, 42) = 13.47 and 34.61, p's $< .001$, and also showed significantly fewer percent errors under the changed than the same condition ($F = 4.90$, $p < .05$). However, the slight overall superiority of 3R over 1R groups was nonsignificant ($F < 1$), although within the same condition the difference in favor of 3R over 1R performance was significant ($F = 4.63$, $p < .05$), leading also to a near-significant 1R-3R × Same-Changed-Nonadjacent interaction, F (2, 42) = 2.96, $p < .10$. Extended-trend analysis of variance showed both significant overall linear and quadratic components, *F*s (1, 42) = 60.99 and 7.97, p's $< .01$, reflecting the positively accelerated increase in errors over the three serial positions clearly displayed in Fig. 1A. However, none of the Group × Position interactions, nor any orthogonal components thereof, approached significance (all *F*s < 1.35, $p > .25$). Similar analyses of the before-errors and rank-score measures revealed virtually identical pictures, except that the overall superiority of changed over same conditions was nonsignificant for both measures, *F*s (1, 28) = 1.55 and 2.38, p's $> .10$.

As shown in Fig. 1B, relative performance by middle cluster groups in Positions 6-7-8 produced a quite different and more complicated picture. Again the same condition led to significantly fewer percent errors than the nonadjacent condition, F (1, 42) = 11.88, $p < .005$, but both the slight difference in favor of the changed over the nonadjacent condition ($F = 2.64$, $p > .10$) and the superiority of same over changed conditions ($F = 3.32$, $p < .10$) were nonsignificant. This latter difference, however, was significant for before errors ($F = 5.07$, $p < .05$) although not for rank scores ($F = 2.22$, $p > .10$). While extended-trend analysis showed no sig-
]

nificant overall differences between the three cluster positions nor any Same-Changed-Nonadjacent × Positions interaction (all $Fs < 1$), both the overall 1R-3R × Positions interaction, $F\ (2, 84) = 4.79, p \approx .01$, and the linear component thereof, $F\ (1, 42) = 5.60,\ p < .025$, were significant. As shown in Fig. 1B, this interaction reflects primarily the marked superiority in Position 8 for both same and changed 3R groups as compared with the 1R groups, resulting further in a highly significant linear component ($F = 8.70,\ p \approx .005$) and also a significant quadratic component ($F = 5.28,\ p \approx .025$) of the Same-Changed × Positions interaction. The improvement over positions for 3R groups, as contrasted with the 1R decrement, led also to significant linear components of the 1R-3R × Positions interactions for both before errors ($F = 11.70, p < .005$) and rank scores ($F = 9.64, p < .005$), although for neither measure did the quadratic component approach significance (both $Fs < 1$).

The opposite direction of the same-changed comparisons for beginning and middle positions is indicative of consistently better List 2 performance on items from the middle rather than the beginning of List 1, irrespective of whether these appeared in the same or changed positions in List 2. This difference was corroborated by an analysis of variance including both beginning and middle List 2 groups, which indicated a significant superiority of List 1 middle over beginning conditions in cluster performance, both for percent cluster errors, $F\ (1, 60) = 6.30,\ p < .025$, and before errors ($F = 6.58,\ p < .025$), although not for rank scores ($F = 3.90,\ p < .10$).

Nonadjacent comparisons.—Figure 1 offers little evidence that nonadjacent performance was better on the middle cluster item, which appeared in the same position as in List 1, than on the preceding and following "cluster" items not included in List 1. Analyses presented in the preceding section revealed no significant differences in shape between these nonadjacent curves and the combined same-changed curves, either for beginning or middle cluster conditions. Likewise, separate analyses of variance for the nonadjacent conditions showed no significant differences in percent errors between the position containing the List 1 item and the preceding and following positions. Only the borderline significance of the quadratic component, $F\ (1, 28) = 4.57, p < .05$, for the beginning-position function in Fig. 1A is suggestive of any facilitative effect of including in List 2 a single item in the same position as in List 1, and even this finding must be discounted in view of the comparable shape of the various same and changed curves in Fig. 1A. Similar comparisons between Positions 9-10-11, which contained the same three items (QE-GV-YB) for all nonadjacent groups, showed that more errors were actually made on the List 1 item (GV) appearing in Position 10 (18.9) than on either the immediately preceding (15.3) or following (16.4) item.

List 1-2 comparisons.—The close similarity and lack of significant differences in shape between the same and changed curves, as shown in Fig. 1, indicate differences between the List 1 beginning and middle groups in relative degree of learning of the three cluster items to have played a minor role, if any, in List 2 performance on these items. Nonetheless, in an attempt to adjust for possible differences between performance in individual cluster positions, the difference between percent errors made on each List 2 cluster item, and the percent errors made on the item appearing in that same serial position in List 1, was calculated for each *S*. The validity of this comparison is suspect due to the lesser learning criteria required for List 1 than List 2, leading to a general overestimate of List 1 percent errors for beginning positions and a compensating underestimate for middle positions, as compared with List 2. However, despite these limitations, analyses of variance on these difference scores generally supported the findings of the List 2 cluster comparisons. Likewise, a separate analysis based on difference scores for the three List 1 items (3-7-10) in the nonadjacent condition revealed no significant List 1-2 differences in percent errors for any of the three positions.

Noncluster performance.—As expected, significantly fewer percent errors per item were made by List 2 beginning (2.7) than middle (8.7) conditions in Positions 2-3-4, as well as by middle (7.1) than beginning (12.1) conditions in Positions 6-7-8 (all Fs > 22.4, p's $< .001$). Separate analyses of percent errors made by the various groups were also performed for Position 1, Position 5, and the last four positions (9–12). No significant group differences were found in Position 1 performance ($F < 1$), but significantly fewer percent errors were made in Position 5 by List 2 beginning (7.9) than middle (10.5) cluster groups, $F\,(1, 84) = 8.29, p < .01$. By contrast, middle cluster groups made significantly fewer percent errors per item (9.7) than did beginning cluster groups (11.3) in Positions 9–12, $F\,(1, 84) = 12.40, p < .001$. Although nonadjacent groups made consistently more total errors in all noncluster positions than did same and changed groups, the percent of these occurring in Positions 9–12 (38.6) was significantly lower than for either same (43.3) or changed (40.9) groups, $F\,(2, 84) = 3.33, p < .05$, due to the marked improvement in performance by these latter groups in their respective cluster positions and consequent relative increment in percent errors in the final noncluster positions. Neither List 1 cluster position nor 1R-3R criterion had any significant effect, either alone or in interaction, on relative performance in any of the noncluster positions.

DISCUSSION

The present finding of a consistent intracluster improvement in performance only across middle cluster positions, and only under 3R conditions, clearly indicates different learning processes to be involved at the beginning and middle of a serial list, and further suggests these processes to change systematically with practice. More specifically, it would appear that *S*s initially attempt to identify and learn the first item(s) in the list, by associating these items with either or both immediately adjacent items or serial positions. As learning proceeds toward the middle of the list, however, progressive increases in intralist interference render either simple-adjacent or position associations increasingly inadequate, and lead to the gradual development of stronger and more complex multiple-item associative units, whereby each item is associated with a complex of cues including several preceding items.[5] In addition to more successfully resisting sources of intralist interference, these complex associative networks also become less dependent upon the maintenance of the context of the entire original list and are thus better able to withstand a shift to new lists or serial positions, as indicated by the consistent superiority of List 2 performance on List 1 middle over beginning cluster items, even though these items did not differ in number of actual List 1 correct responses.

---

[5] Processes of learning the end of a serial list will be omitted from the present discussion, since these were not directly investigated here. However, in line with a suggestion by Peterson, Brewer, and Bertucco (1963), subsequent research in this laboratory (Mathews, 1963) offers evidence for an exclusion process at the end of the list, such that *S*s learn to respond within each trial only with items which have not appeared previously in the list.

To the extent that different learning processes become involved in various parts of a serial list, it necessarily follows that items must somehow become localized and associated at least with their general location in the list. However, the present results are uniformly negative with regard to associations of individual items with specific serial positions as a principal mechanism of serial learning. The absence of any List 2 facilitation on the three items appearing in nonadjacent serial positions identical to their List 1 positions, despite *S*s' prior familiarization with these items, not only argues strongly against specific position associations, but also mitigates what otherwise might have been a serious flaw in the present design, its omission of control conditions with no List 1 items included in List 2. Such a control condition is thus provided by the nonadjacent groups, which serve also to control for relevance of prior List 1 learning since List 2 contained three List 1 items for all groups.

Also contrary to position associations is the present failure to find significant interfering effects under changed conditions, which instead (*a*) were consistently superior to nonadjacent conditions, (*b*) differed insignificantly from same conditions in overall performance, and (*c*) were actually significantly superior to same conditions in beginning cluster positions. Only the slight superiority of same over changed conditions in middle positions (Fig. 1B), which was significant only for before errors, is suggestive of any decremental effects of changed serial positions in List 2, and this may well reflect only the consistently greater facilitation of List 2 performance for List 1 middle than beginning cluster items.

The traditional chaining explanation, in terms of direct sequential associations between adjacent items, is entirely consistent with the lack of interference under changed conditions. It cannot, however, readily handle the obtained intracluster effects, nor the differences between these for beginning and middle positions. A chaining hypothesis would predict better performance on the second cluster item, for which both immediately preceding and following items were identical to List 1, than on the first and third cluster items, for which the preceding and following items, respectively, were changed. Not only does Fig. 1 fail to display such intracluster differences, but performance was also significantly facilitated on immediately following noncluster items, both in Position 5 for beginning and Positions 9–12 for middle cluster conditions. This asymmetrical generalization effect further indicates serial learning to involve primarily unidirectional forward associations, and according to the present analysis would be attributed to the presence of immediately preceding multiple-item associative units, into which the subsequent noncluster items can be incorporated.

Among the present findings, only the failure to find greater facilitation across beginning cluster positions under 3R than 1R changed conditions (as was evident in middle cluster positions for both same and changed conditions) appears at all inconsistent with the proposed multiprocess account. It is not unlikely, however, that these beginning-changed groups attempted to utilize the hypothesized simpler processes of learning List 2 beginning items, which proved to be less adequate for Position 4 than for the two preceding positions, and thus obscured any intracluster differences between 1R and 3R conditions.

According to the present account, the relative involvement and complexity of multiple-item associative units should increase systematically with list difficulty or other conditions conducive to intralist interference. Consequently, with easier lists or conditions than those of the present study, simple adjacent or position associations alone may be sufficient to learn the list. In this connection, it is noteworthy that previous evidence for position associations has been obtained predominantly with highly meaningful items (Young, 1959, 1961, 1962) or with shorter lists and slower presentation rates (Ebenholtz, 1963a, 1963b). In contrast, previous studies with longer lists and high-interitem associative strength (Horowitz & Izawa, 1963) or low-item meaningfulness (Underwood, 1963) have yielded little evidence for position associations, in agreement with the present results. Also consistent with implications of the present analysis are Ebenholtz' (1963b) finding that position associations are more important at the ends than the middle of the list, and the demonstration by Young et al. (1963) of greater positive transfer at the middle than at the ends of a reversed serial list.

Besides its consistency with a variety of previous results, the present interpretation has much in common with accounts offered previously by Jensen (1962a) and others. It is distinguished principally by its attempted specification of the relative contributions of various hypothesized mechanisms of serial learning as a function of position in the list, level of practice, and various task and procedural factors. Future research can thus profitably be directed toward the further specification of the nature of the suggested multiple-item associative networks and of the factors relevant to their development.

REFERENCES

ASCH, S. E., HAY, J., & DIAMOND, R. M. Perceptual organization in serial rote-learning. *Amer. J. Psychol.*, 1960, **73**, 177–198.

BATTIG, W. F., BROWN, S. C., & NELSON, D. Constant vs. varied serial order in paired-associate learning. *Psychol. Rep.*, 1963, **12**, 695–721.

BROWN, S. C., & BATTIG, W. F. Partial serial-position constancy in nonserial paired-associate learning. *J. verbal Learn. verbal Behav.*, 1962, **1**, 42–47.

EBENHOLTZ, S. M. Position mediated transfer between serial learning and a spatial discrimination task. *J. exp. Psychol.*, 1963, **65**, 603–608. (a)

EBENHOLTZ, S. M. Serial learning: Position learning and sequential associations. *J. exp. Psychol.*, 1963, **66**, 353–362. (b)

FEIGENBAUM, E. A., & SIMON, H. A. A theory of the serial position effect. *Brit. J. Psychol.*, 1962, **53**, 307–320.

HOROWITZ, L. M., & IZAWA, C. Comparison of serial and paired-associate learning. *J. exp. Psychol.*, 1963, **65**, 352–361.

JENSEN, A. R. An empirical theory of the serial-position effect. *J. Psychol.*, 1962, **53**, 127–142. (a)

JENSEN, A. R. Transfer between paired-associate and serial learning. *J. verbal Learn. verbal Behav.*, 1962, **1**, 269–280. (b)

MATHEWS, S. R. Serial learning under conditions of prior familiarization with part of the list. Unpublished master's thesis, University of Virginia, 1963.

MCCRARY, J. W., & HUNTER, W. S. Serial position curves in verbal learning. *Science*, 1953, **117**, 131–134.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, 1951.

MURDOCK, B. B. The distinctiveness of stimuli. *Psychol. Rev.*, 1960, **67**, 16–31.

PETERSON, L. R., BREWER, C. L., & BERTUCCO, R. A guessing strategy with the anticipation technique. *J. exp. Psychol.*, 1963, **65**, 258–264.

REHULA, R. J. A test of two alternative hypotheses of the associations that develop in serial verbal learning. Unpublished doctoral dissertation, Northwestern University, 1960.

SCHULZ, R. W. Generalization of serial position in rote serial learning. *J. exp. Psychol.*, 1955, **49**, 267–272.

UNDERWOOD, B. J. Stimulus selection in verbal learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963. Pp. 33–48.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

VOSS, J. F. A comparison of motor and verbal serial learning with length of sequence and verbal material as parameters. Unpublished doctoral dissertation, University of Wisconsin, 1956.

YOUNG, R. K. A comparison of two methods of learning serial associations. *Amer. J. Psychol.*, 1959, **72**, 554–559.

YOUNG, R. K. The stimulus in serial verbal learning. *Amer. J. Psychol.*, 1961, **74**, 517–528.

YOUNG, R. K. Tests of three hypotheses about the effective stimulus in serial learning. *J. exp. Psychol.*, 1962, **63**, 307–313.

YOUNG, R. K., PATTERSON, J., & BENSON, W. M. Backward serial learning. *J. verbal Learn. verbal Behav.*, 1963, **1**, 335–338.

(Received June 13, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 458–462

ASSIMILATION AND CONTRAST IN THE ESTIMATION OF NUMBER[1]

WILLIAM BEVAN AND EDWARD D. TURNER

Kansas State University

Independent groups were presented random scatters of dots on a gray field and asked to estimate the number in each. The experimental conditions consisted of surrounding each scatter with either a large or small frame and giving instructions intended to identify scatter and frame together as focal stimuli or to segregate them, one as figure, the other as part of the background. Under the "figure" condition the frames were found to exert an assimilative effect upon the judgment of the dots. Under the "ground" condition the opposite was found to be the case. The colors of dots and frames relative to the overall field were not found to influence judgment.

The commonly observed result of introducing an anchor stimulus into a judgmental sequence is the *contrast effect*, i.e., the shift of the judgmental responses away from the anchor magnitude. Thus, if the anchor lies above the series, the judgments are shifted downward. Recently, Sherif, Taub, and Hovland (1958) have indicated that the opposite, which they term *assimilation*, may also occur. This has been confirmed by Parducci and Marshall (1962). According to Sherif and Hovland (1961, p. 49), assimilation is less marked than contrast and results from anchors within or near the series, while contrast is produced by more distant anchors.

The present experiment seeks to explore the underlying mechanisms that prompt contrast, on the one hand, and assimilation, on the other. A ready suggestion comes from Gestalt psychology and relates to the manner in which the judge organizes the perceptual field. In the experimental situation most frequently employed in anchor studies, the judge may be expected to conceptualize the stimuli, both series and anchor, as distributed along a single judged dimension with other differences controlled or viewed as irrelevant. Certain of the organizing principles, which describe relationships among spatial arrays of stimuli, also may be extended reasonably to characterize the conceptual relationships just mentioned. Thus, Sherif and Hovland's hypothesis concerning the separation of anchor and series may be considered in terms of the Principle of Proximity and Bevan and Pritchard's (1963) recent observation that marked changes in the color or size of an anchor-designate for shape resulted in a complete loss of its anchor properties may be thought of in terms of the Principle of Similarity.

There are, of course, situations for studying anchor effects in which the properties of spatial arrays clearly assume significance. The study of the estimation of number provides a convenient example. Here the common procedure involves the presentation for judgment of scatters of dots. This situation, moreover, suggests a conceptual similarity of possible significance between contrast and assimilation and the Gestalt principle of figure-ground field articulation. Figure and ground are opposite and com-

---

[1] This experiment was performed under Contract Nonr-3634(01) between the Physiological Psychology Branch, Office of Naval Research and Kansas State University.

plementary effects. Contrast and assimilation are opposite effects which also may be complementary. If the anchor-designate is identified, along with the series stimuli, as a focal stimulus ("figure"), internal articulation should be reduced and assimilation should occur. If, however, the anchor-designate is differentiated from the series as a *contextual stimulus* ("ground"), contrast should occur.

This experiment extends an earlier experiment by Bevan, Maier, and Helson (1963) in which *S*s judged the number of beans in a series of large or small jars under instructions designed either to identify beans and jars as unitary wholes ("figure" instructions) or to emphasize their separate identities ("ground" instructions). Under "figure" instructions, estimates of the numbers of beans in large jars exceeded those of the same actual count when in small jars (assimilation effect). No contrast effect was in evidence under the condition of "ground" instructions.

The materials of the Bevan, Maier, and Helson experiment, however, do not readily allow for the collection of control data from which to assess directly the occurrence of assimilation or contrast. In addition, it is conceivable that such effects are more easily demonstrable with more abstract but conceptually simpler stimuli. Therefore, in the present experiment, *S*s were presented random arrays of dots on a homogeneous field. It was assumed, furthermore, that any stimulus that may be related conceptually to the stimuli to be judged may display anchor properties. Since the dots to be judged in this task represented a spatial array, it appeared appropriate to employ a frame that enclosed the array as the anchor-designate.

It was clear that assimilation and contrast effects should be opposite in kind. At the same time it remained a problem to anticipate what the nature of these effects would be under the present experimental conditions. It seemed reasonable to assume that if the total area encompassed by the frame relative to that of the scatter was large (i.e., the inside dimensions of the frame and the perimeter of the scatter were separated clearly and to some substantial degree) and if the field was organized so as to establish the structural unity of scatter-frame, the "gap" between the scatter and the judgmental referent (the frame) would be reduced by an enhancement of the apparent area of the scatter. If, however, the field was organized so as to emphasize the separate identities of the scatter and frame, differentiation would produce enhancement of the "gap" between the scatter and the anchor with a reduction in the apparent area of the scatter. On the other hand, if the total area encompassed by the frame relative to that of the scatter was small (i.e., the inside dimensions of the frame were coincident with or in close proximity to the perimeter of the scatter), it was conceivable that the judge would categorize the frame as small and under "figure" instructions the effect would be one of general compression in areal scatter. If, however, the complementarity of scatter and frame-anchor were stressed, there would be a tendency to compensate for the compression by an overestimate of scatter area.

Accordingly, two sets of intuitive hypotheses were formulated: (*a*) If the frame were large relative to the area of the array and if it were identified by instructions as part of the focal stimulus, then the estimates of number obtained under this condition would exceed those of the control condition, when no frame was present. If, however, the frame were differentiated from the array as part of the background, the estimates of number would be less than those obtained without the frame present. (*b*) If the frame were small (only slightly larger than the area of the array), the opposite set of relationships would obtain among the estimates for the several conditions.

METHOD

Experimental design.—The experimental design was a 2 × 2 × 2 factorial involving instructions (figure vs. ground), frame size (large vs. small), and color of dots and frame (black vs. white on a medium-gray field). This was supplemented by two control groups (black dots vs. white dots, both run without frame or differentiated instructions) to provide baseline data.

Subjects.—One hundred fifty students in introductory psychology were divided randomly into 10 equal-sized groups. There was approximately the same number of male and female *S*s in each. Five groups judged white on gray; five black on gray.

Stimulus materials.—The psychophysical series consisted of five random arrays containing 21, 42, 84, 168, and 336 dots, respectively. Each was arranged at the center of a large (22 × 28 in.) gray (Munsell No. 7) poster board so as to demarcate roughly a square 9 × 9 in. The dots were .25 in. in diameter, made either of black or white construction paper. For the conditions involving frames, squares were added in matching color. The large frame was 20 × 20 in. inside dimensions, with segments 2 in. thick. The small frame was 10 × 10 in. with sides 1 in. thick.

Procedure.—The *S*s were tested in groups of 15. The controls were given these instructions:

> This is an experiment in the estimation of number. I am about to show you a series of gray cards on which are random scatters of dots. When I place each card before you, I want you to write down on the answer sheet I have just given you the number of dots you think is on that card. *Do not count* or otherwise try to calculate, but give me *your first impression.*

In the groups receiving frame as *focal stimulus*, the following was added:

> You will note that the random scatter is bounded by a thick frame, setting the whole arrangement off from the gray background as a single unit. It is important that you recognize that dots and frame go together in a unitary pattern.

Those being presented the frame as *context* received these added instructions:

> You will note that the random scatter is bounded by a thick frame. It may be natural for you to think of dots and frame as one, but remember they are separate. We are interested in your estimates of the dots. The frame is just part of the background.

Next, the Rubin goblet figure was shown to the "frame" groups and the differences between figure and ground briefly amplified.

After the instructions were given, each card was presented for 2 sec., beginning with the smallest number and proceeding to the largest. When the series was completed, it was repeated twice again and the median of the three estimates for each stimulus array was taken as the value to be used in data analysis.

RESULTS AND DISCUSSION

The data of this experiment are graphically presented in Fig. 1. Inspection of the control data, obtained for the series on the homogeneous gray ground, indicates a general tendency toward underestimation, slight, if indeed present, for the 21-dot array and increasing until quite marked for the 336-dot array. This is in line with the earlier data of Bevan, Maier, and Helson, who found a decided underestimation under all their conditions.

Figure 1 also clearly indicates that the hypotheses advanced above have been confirmed without exception. The slopes for all experimental conditions are clearly different from that for the control. When "figure" instructions were used with the large frame, the estimates tended increasingly to be overestimated relative to the control. When they were employed with the small frame, the estimates tended increasingly toward underestimation. Meanwhile, for "ground" instructions, these relationships are reversed. With the large frame, they produced underestimations relative to the control and, with the small frame, overestimations. Figure instructions thus produced assimilation effects; ground instructions, contrast effects. (It is interesting to note that these effects were greatest for the arrays containing the greater number of dots, where the greatest inaccuracy in judgment also is found, while little change occurred at the lower end of the series where judgments are most accurate.) These results were statistically verified by a highly significant Instructions × Frame Size interaction, $F(1, 112) = 18.02$, $p < .005$, for the four experimental conditions.

It is furthermore apparent from Fig. 1 that frame as focal stimulus exerted a greater effect upon judgment than frame as context. This is in agreement with the earlier data of Bevan, Maier, and Helson but inconsistent with Sherif and Hovland's (1961) interpretation of assimilation and contrast which implies that contrast necessarily must be the more potent process.

Finally, inspection of Fig. 1 suggests that the large frame exerted a stronger influence upon judgment than did the small. This is analogous to the differential effect associated with anchors near and more distant from the series.

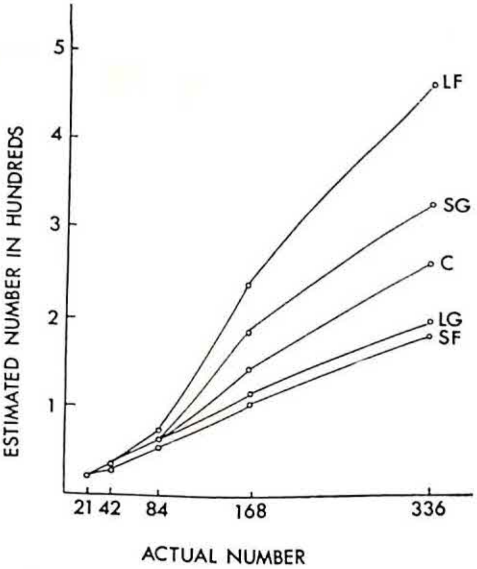


FIG. 1. Estimations of the number of dots in a series of random arrays, when bordered by large or small frames, and when made under instructions which emphasized the integral relationship between scatter and frame ("figure" instructions) or their distinctness ("ground" instructions). (C represents the control—without frame—condition. The curves marked F represent the conditions when "figure" instructions were used; those marked G the "ground" conditions.)

Sherif and Hovland (1961, pp. 48–51) identify the functional differences between assimilation and contrast with differences in judgmental attitude. For them assimilation implies perceived similarity between related aspects of the stimulus situation, while contrast implies perceived differences. This can be effectively illustrated by comparing attitude-scale values assigned to items relating to certain social issues when the content of these items is identified by the judge, on the one hand, with his own position and, on the other, with that of an out-group. However, it need not be assumed that the perception of similarities and of differences are necessarily mutually exclusive processes, that they are phenomenal polar-opposites, or that they are independent cognitive events. At the same time, one may interpret the operations of the present experiment, designed to effect differences in organization of the perceptual field, also as devices for emphasizing similarity on the one hand and difference on the other. Meanwhile, Hovland and Sherif, by equating similarity with dimensional proximity, have been logically forced to the prediction that contrast effects are necessarily stronger than assimilation effects.

Similarly, Parducci and Marshall (1962), in their elaboration of the Sherif, Taub, and Hovland (1958) study, emphasized the importance of the judge's understanding of the rules of judgment. According to this view, in an unanchored series or a series with the anchor sufficiently remote, the implicit rules of judgment place the standard near the midpoint of the series. When, however, the anchor is placed within or in juxtaposition to the series, *E* has created a rule of judgment which identifies anchor as standard and assimilation and contrast become interpretable within the context of adaptation-level theory, as complementary phenomena. In the present situation, the rule becomes one, depending upon instructions, that emphasizes either a similarity *or* a difference between series and frame anchor.

The color of the figures (black vs. white) had no detectable effect upon judgment, $F\,(1, 112) = .439$, $p > .05$.

REFERENCES

BEVAN, W., MAIER, R., & HELSON, H. The influence of context upon the estimation of number. *Amer. J. Psychol.*, 1963, **76**, 464–469.

BEVAN, W., & PRITCHARD, J. F. The anchor effect and the problem of relevance in the judgment of shape. *J. gen. Psychol.*, 1963, **69**, 147–161.

PARDUCCI, A., & MARSHALL, L. M. Assimilation and contrast in the anchoring of perceptual judgments of weight. *J. exp. Psychol.*, 1962, **63**, 426–437.

SHERIF, M., & HOVLAND, C. I. *Social judgment.* New Haven: Yale Univer. Press, 1961.

SHERIF, M., TAUB, M., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55**, 150–155.

(Received June 18, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 463–467

OVERLEARNING-EXTINCTION EFFECT AS AN INCENTIVE PHENOMENON[1]

JOHN THEIOS AND JOHN BRELSFORD

University of Texas

3 groups of 12 rats each were trained and extinguished in a runway. Group 10-10 received 10 rewarded training trials. Group 60-60 received 60 rewarded training trials. Group 10-60 received 10 training trials and 60 rewards. An overlearning effect in extinction was found only when the groups differed in number of rewards. Group 10-10 was more resistant to extinction than Group 60-60, but Group 60-60 was more resistant than Group 10-60. The data suggest that the overlearning effect is a result of the operation of incentive motivation.

Ison (1962) has published runway data in which resistance to extinction is a strict monotone *decreasing* function of number of rewarded training trials. On the surface, these data seem preposterous in light of all extant theories of learning. Ison (1962) suggests that his runway data are analogous to results on reversal learning following overlearning in a choice or discrimination situation. The typical overlearning effect is that *S*s who receive an extended period of original training learn to reverse their choice behavior faster than *S*s who are trained only to a criterion (Birch, Ison, & Sperling, 1960; Capaldi & Stevenson, 1957; Pubols, 1956; Reid, 1953). It should be mentioned that in all the studies cited above, the reward was of an appetitive nature; i.e., food.

The term *overlearning-reversal effect* has been used to describe results of choice experiments in which *S*s trained to a criterion take longer to reverse their choice behavior than *S*s whose training is continued after the criterion is reached. In a similar manner, the term *overlearning-extinction effect* will be used in this paper to describe results of runway experiments such as Ison's (1962) in which *S*s given a large number of training trials extinguish faster than *S*s given a smaller number of trials.

The overlearning paradox vanishes when one grasps a single point, namely, that the effectiveness of a nonrewarded trial should not be the same following a small number as following a large number of appetitively rewarded trials. A decrement in response strength on a nonrewarded trial following a large number of rewarded trials should be much larger than that following only a few rewarded trials. Thus, resistance to extinction cannot be taken as a valid measure of response strength when groups differ in number of rewarded trials. This idea becomes very clear when it is cast in terms of one of the traditional behavior theories (Spence, 1956). According to the theory, habit strength (H) is an increasing function of the number of training trials (N) irrespective of whether they are reinforced or not. Incentive motivation (K) is an increasing joint function of the number of times reward has been experienced at the goal (N_g) and amount of reward (W_g). Resistance to extinction (R_e) is usually taken as an increasing function of both H and K. There is a sizable amount of literature, however, which indicates that resistance to extinction following continuous reward is a *decreasing* function of amount of reward (e.g., Armus, 1959; Hulse, 1958; Wagner, 1961). In terms of the theory, these data suggest that R_e is a decreasing function of K. The mechanism for this relationship could be disruption of the instrumental response by stimuli introduced as a result of frustrative nonreward (Amsel, 1962).

---

[1] This investigation was supported in part by Public Health Service Research Grant MH-07471-01 from the National Institute of Mental Health, Public Health Service.

The thesis of the present paper is that the overlearning literature can be understood if three assumptions are made:

1. Resistance to extinction is an increasing function of number of times the organism performs the instrumental response, $R_e = +\,f(H)$.
2. Resistance to extinction is a decreasing function of number of appetitive rewards experienced at the goal, $R_e = -\,f(K)$.
3. The effect of number of trials asymptotes before the effect of number of rewards.

From the assumptions, it follows that the overlearning effect will result in appetitive reward situations, since overtraining the original response is more likely to be disrupted and therefore less resistant to extinction. That this is indeed what happens is documented by the data of the single-stimulus-presentation study of Birch, Ison, and Sperling (1960). They report that

> the difference between the two groups appears to be located in their differential rates of extinction to the previously reinforced cue. No difference in rate of acquisition to the newly positive cue was found in the present data [pp. 39–40].

The present experiment was a test of the proposed interpretation of the overlearning effect. Using Ison's (1962) technique, two of his conditions were replicated. These were the groups which received 10 and 60 rewarded training trials. They thus may be termed the 10-10 and 60-60 groups, respectively, where the first member of the dyad designates the number of running trials and the second member, the number of rewards. A novel, but very important, group, 10-60, was introduced in the present study. This group received only 10 running trials through the alley, but received 60 rewards at the goal. It is predicted that (*a*) the 10-10 group should be more resistant to extinction than the 10-60 group, since the groups are equated on H but differ in K, and (*b*) the 60-60 group should be more resistant than the 10-60 group, since the groups are equated on K but differ in H. In other words, an overlearning effect should not occur if the groups are equated on K.

METHOD

Apparatus.—The apparatus consisted of a black wooden runway 5 in. wide, 5 in. high, and 72 in. long. The first 12 in. comprised the start box and the last 12 in., the goal box. The start and goal boxes were separated from the alley by metal guillotine doors. Response time was measured by a Lafayette electric stop clock in .01-sec. intervals from the opening of the start door until S interrupted a photobeam 5 in. inside the goal box.

Subjects.—The *S*s were 36 naive, male albino rats of the Holtzman strain, approximately 90 days old at the start of the experiment. They were housed in individual living cages with ad-lib. water.

Procedure.—The procedure for the 10-10 and 60-60 groups was similar to that for the 10 and 60 groups of Ison's (1962) study, which should be consulted for details. Briefly, all *S*s were placed on a 15-gm. feeding schedule 14 days prior to training and were handled each day during this period. Two training trials were given on the first and second days, and 5 trials per day thereafter, except for the first day of extinction which contained the final rewarded trial followed by 5 nonrewarded trials. The reward was a .7-gm. wet mash pellet placed in a glass coaster at the end of the goal box.

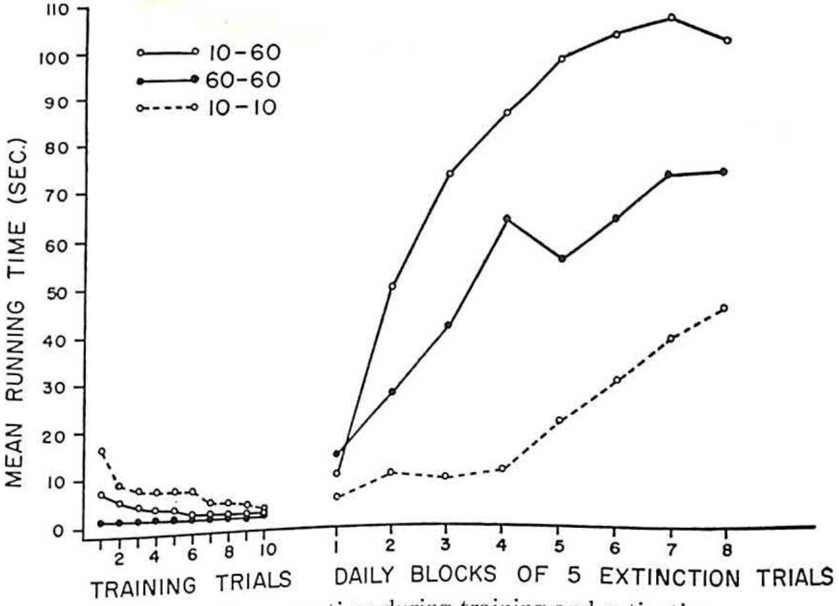


FIG. 1. Response time during training and extinction.

The 10-10 group was given 10 rewarded running trials and then extinguished. The 60-60 group was given 60 rewarded trials and then extinguished. On the first 50 trials, each 10-60 *S* was merely placed in the baited goal box, with his head directly over the food cup. On Trials 51–60, however, *S* was placed in the start box and he had to traverse the runway to obtain the reward. Thus, the 10-60 group began extinction on the same day as the 60-60 group, and all nonassociative factors, such as time on the drive schedule, should be the same for these two groups.

During extinction, all *S*s were run for a minimum of 80 trials, or until they met a criterion of not entering the goal box within 120 sec. The empty food cup was present, and *S* was detained in the goal box for 30 sec. The intertrial intervals during acquisition and extinction averaged about 20 min.

RESULTS

The mean number of trials to the first 120-sec. extinction trial was 61 for the 10-10 group, and 26 for the 60-60 group. The values for the corresponding groups in Ison's (1962) study were 60 and 21. Thus, we have effectively replicated Ison's results in the present study. In fact, the replication agreement is remarkable, considering that different apparatus, strain of rats, etc., were used in the two studies.

The 10-60 group, however, took 18 trials to reach the 120-sec. extinction criterion. The significance of the differences among the three groups was tested by Duncan's multiple-range test. A log (base 10) transformation was performed on the criterion scores to equate the variances of the groups. The means for the transformed scores were 1.67, 1.38, and 1.20 for the 10-10, 60-60, and 10-60 groups, respectively. The corresponding unbiased variance estimates were .031, .031, and .045. The range test indicated that all three differences among the groups were significant at well beyond the .05 level (least significant range = .16, $df = 33$). Thus, the 10-10 group was significantly more resistant than the 60–60 group, which, in turn, was significantly more resistant than the 10-60 group.

Running times on the last 10 training trials are presented in the first part of Fig. 1. The groups are ordered as expected, with the 60-60 *S*s asymptotic at 1.1 sec. and the 10-60 group performing slightly better than the 10-10 group on the 10 training trials. The overall difference among the groups on the 10 trials is significant, $F\ (2, 33) = 48.7, p < .001$.

Running time during the first 40 trials of extinction is presented in the second part of Fig. 1. The differences among the groups in resistance to extinction are clearly evident. The overall difference among the groups on the 40 trials is significant, $F\ (2, 33) = 45.4, p < .001$.

DISCUSSION

An overlearning effect of decreased resistance to extinction with increased trials in a runway was demonstrated by Ison (1962). In the present study, however, an overlearning effect of decreased resistance to extinction with increased training trials was not obtained when the criterion group (10-60) and the overlearning group (60-60) were equated on number of rewards experienced at the goal. On the other hand, when the criterion (10-10) and overlearning (60-60) groups were not equated on number of rewards, an overlearning effect of decreased resistance to extinction with increased training trials was obtained. The data support the assumptions that resistance to extinction is (*a*) an increasing function of the number of times the animal performs the response and (*b*) a decreasing function of number of rewards experienced at the goal.

Collateral support for the assumptions comes from a "reversal" study of Ison and Birch (1961). They gave *S*s a differential number of rewarded placements in one end box of a T maze and then trained them to go from the start box to the opposite end box. Since both groups at the start of "reversal" had equal experience (i.e. none) in running the T maze, the group given the greater number of rewards in the opposite end box should learn the "reversal" faster. This is, in fact, what happened.

The assumptions have obvious referents in the behavior systems of Spence (1956) and Amsel (1962). Habit strength (H) is an increasing function of training trials (N), and resistance to extinction (R_e) is an increasing function of H. Incentive motivation (K) is an increasing function of number of rewards (N_g), and the magnitude of disruptive frustrative responses (r_f) on nonrewarded trials is an increasing function of K. Finally, R_e is inversely related to r_f.

One remaining problem is to account for the not infrequent failure to find an overlearning effect (D'Amato & Jagoda, 1962; D'Amato, Schiff, & Jagoda, 1962; Hill, Spear, & Clayton, 1962). An examination of the overlearning literature revealed that studies failing to obtain the overlearning effect have used a reward of a very small size. Hill et al. used one or two .045-gm. Noyes pellets as rewards. D'Amato et al. used single .045-gm. Noyes pellets or 1.75-sec. long drinks of water. It is inferred that 1.75 sec. drinking per trial is a small reward, since D'Amato and Jagoda (1962) gave their *S*s as many as 20 trials per day. In contrast, 20 .7-gm. wet mash pellets of the type used by Ison (1962) and the authors would approximate our rats' daily rations. The observation of differences in the size of the reward becomes important when it is postulated that the asymptote of K is an increasing function of the magnitude of the reward. For large rewards, the curves for H and K as a function of rewarded trials would diverge very quickly and then slowly converge during most of the domain. If during extinction K essentially subtracts from H, as we are tacitly assuming, then an overlearning effect will result in all studies except those that use a very small number of original and overlearning trials. Data which nicely confirm this prediction have been collected in a dissertation study by L. T. Clifford.[2] For five groups of rats, Clifford varied number of runway trials up to 192. It was found that resistance to extinction rapidly increased at first and then slowly decreased as a function of training trials.

---

[2] The results of Clifford's study have been reported by Birch (1962).

For small rewards, the curves for H and K as a function of trials will diverge for most of the domain. If, with small rewards, K reaches its low asymptote before H reaches its high asymptote, then the curves will never converge, and the overlearning effect could not result. Consideration of various curves for H and K leads to the conclusion that whether or not the overlearning effect will be obtained in a study will be determined by number of criterion trials, number of overlearning trials, rate of growth, and asymptotes of the H and K functions.

The basic thesis of this paper is also in accord with the theory of cognitive dissonance as applied to the behavior of rats by Festinger (1961) and Lawrence and Festinger (1962). On the nonrewarded trials at the start of extinction or reversal learning, *S* will experience dissonance as a result of performing the response and not being rewarded. The magnitude of the dissonance will be a joint increasing function of the number and size of the rewards *S* has experienced. In the present study, the 10-10 groups should be more resistant to extinction than the 10-60 groups because the 10-60 *S*s would experience a large amount of dissonance on extinction trials. As a result, they should change their behavior (i.e. stop running) before the 10-10 *S*s, who would experience a lesser amount of dissonance. Since the 10-60 and 60-60 groups had the same number of rewards, they should experience the same amount of dissonance on extinction trials, and the difference in their resistance to extinction would be due to the differential level of training of the instrumental response.

REFERENCES

AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychol. Rev.*, 1962, **69**, 306–328.

ARMUS, H. L. Effect of magnitude of reinforcement on acquisition and extinction of a running response. *J. exp. Psychol.*, 1959, **58**, 61–63.

BIRCH, D. Extinction effects in reversal learning. Paper presented at American Psychological Association meetings, St. Louis, 1962.

BIRCH, D., ISON, J. R., & SPERLING, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

CAPALDI, E. J., & STEVENSON, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

D'AMATO, M. R., & JAGODA, H. Overlearning and position reversal. *J. exp. Psychol.*, 1962, **64**, 117–122.

D'AMATO, M. R., SCHIFF, D., & JAGODA, H. Resistance to extinction after varying amounts of discriminative or nondiscriminative instrumental training. *J. exp. Psychol.*, 1962, **64**, 526–532.

FESTINGER, L. The psychological effects of insufficient rewards. *Amer. Psychologist*, 1961, **16**, 1–11.

HILL, W. F., SPEAR, N. E., & CLAYTON, K. N. T maze reversal learning after several different overtraining procedures. *J. exp. Psychol.*, 1962, **64**, 533–540.

HULSE, S. H., JR. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

ISON, J. R. Experimental extinction as a function of number of reinforcements. *J. exp. Psychol.*, 1962, **64**, 314–317.

ISON, J. R., & BIRCH, D. T-maze reversal following differential endbox placement. *J. exp. Psychol.*, 1961, **62**, 200–202.

LAWRENCE, D. H., & FESTINGER, L. *Deterrents and reinforcement: The psychology of insufficient rewards.* Stanford: Stanford Univer. Press, 1962.

PUBOLS, B. H., JR. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

(Received June 18, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 468–474

MAGNITUDE OF THE FRUSTRATION EFFECT AS A FUNCTION OF CONFINEMENT AND DETENTION IN THE FRUSTRATING SITUATION[1]

JOHN R. MACKINNON AND ABRAM AMSEL

University of Toronto

The experiment investigated the effect of manipulating time of detention on nonreward trials in 2 sizes of goal box on the frustration effect (FE). A double-runway apparatus was employed. The variables under investigation were manipulated in the 1st goal box (G_1) and the effects of the experimental conditions were observed in Runway 2 in terms of starting, running, and goal-entry times. Magnitude of the FE was greater following 3 or 15 sec. detention than following 90 sec. detention on nonreward trials. During both pretest and test trials, the effects of G_1 size and of detention differ greatly depending on the response measure employed, starting and running times showing quite different pictures.

Several experiments concerned with the motivational properties of nonreward (Amsel, 1962), following an early experiment by Amsel and Roussel (1952), have employed a double-runway apparatus. This is essentially two single runways in series. The *S* runs from a start box (SB) into a first runway (Runway 1) and a first goal box (G_1), which opens into a second runway (Runway 2) and a second goal box (G_2). The typical experimental procedure involves the manipulation of reward and nonreward in G_1 following the Runway 1 response. Reward is always available in the G_2. The *frustration effect* (FE) is an increase in vigor of running (i.e., shorter running time) in Runway 2 following nonreward in G_1. The present experiment is concerned with the effects of two variables on the FE: (*a*) time of detention on nonreward trials in G_1 and (*b*) degree of physical confinement in G_1 prior to running in Runway 2.

Time of detention in G_1 on nonreward trials was manipulated by Amsel and Roussel (1952) in the original experiment describing the FE. They set up three experimental conditions: 5, 10, and 30 sec. detention in G_1, and, although their results suggested that the 10-sec. condition produced the largest FE and that by 30 sec. the effects of frustrative nonreward had already begun to dissipate, they were unable to obtain statistically significant support for the hypothesis. The inconclusive findings were attributed, in part, to the fact that the range from 5 to 30 sec. in a frustrating situation may be one which produces relatively little change in the FE. However, a recent re-examination of the Amsel and Roussel data shows a very consistent picture: there is an inverted V-shaped relationship between size of FE and detention (5, 10, or 30 sec.) which appears consistently on each of the three daily trials.

Apart from the Amsel and Roussel experiment, indirect evidence comes from experiments examining the FE which have quite fortuitously employed detentions in G_1 on nonreward trials ranging from 15 to 30 sec., prior to running in the second runway. There is a suggestion from these studies as well that the greater FEs are obtained with shorter (15-sec.) G_1 detentions.

---

[1] Supported by a grant (G-13895) from the National Science Foundation.

The second variable under consideration in the present experiment, size of G_1, has been studied only indirectly. The various studies of the FE have involved first goal boxes differing somewhat in size, and there is a suggestion that the larger, more significant FEs show up when a smaller G_1 is used. It would seem that a more formal study of the effects of this variable is indicated. The present design permits us to observe the effects of (*a*) time of detention in G_1, (*b*) degree of confinement in G_1, and (*c*) the interaction of these two variables. The experiment also provides a first opportunity to observe the development of the FE under one-trial-per-day conditions.

METHOD

Subjects

Sixty albino rats of the Wistar strain served as *S*s for this experiment. The animals were obtained from Woodlyn Farms, Guelph, Ontario, and were approximately 7 wk. old on arrival at the laboratory. All *S*s were housed in individual cages, and were maintained on ad-libitum feeding for a period of 22 days. Their ages ranged from 67 to 72 days at the beginning of the experiment. Four *S*s died during the course of the experiment, leaving 56 from which data were collected.

Apparatus

The apparatus was an L-shaped double runway, essentially as described by Amsel and Penick (1962). The various components of the apparatus, arranged successively, were a start box, a first runway, a first goal box, a second runway, and a second goal box. The inside surfaces of all components of the apparatus were lined with $\frac{1}{8}$-in. Plexiglas, behind which were placed sheet-metal inserts. In order to differentiate the first and second alley conditions as much as possible, SB, Runway 1, and G_1 contained wall inserts painted flat black. The walls of Runway 2 and G_2 were black and white alternating vertical stripes. The floor of the first three segments was stiffened black rubber matting; the floor of Runway 2 and G_2 consisted of $\frac{1}{8}$-in. brass rods spaced approximately $\frac{1}{2}$ in. apart. The runways were covered with Plexiglas throughout, hinged to allow access to the apparatus. To further differentiate conditions between the first and second sections of the apparatus, a small black food cup ($\frac{7}{8} \times \frac{1}{2} \times \frac{1}{4}$ in.) was mounted on the side of G_1, approximately 1 in. from the door leading to Runway 2. G_2 contained a larger ($2 \times 1 \times \frac{1}{2}$ in.) white food cup hung on the far wall of the end box.

Two guillotine-type start doors, one metal, the second clear plastic, separated the SB from Runway 1 and two similar doors separated G_1 from Runway 2. Metal doors prevented retracing from G_1 and G_2. Illumination was provided by two, 40-w. fluorescent lamps suspended in series approximately 20 in. above the second runway.

A modification of the apparatus, as previously employed, was necessitated by the manipulation of the confinement variable. A removable confinement chamber was inserted in G_1 on half of the trials to decrease its size. The walls of the insert were black and lined with Plexiglas, in this respect similar to the larger G_1. The insert had a Plexiglas top, and contained a small metal food cup, mounted on the side, of the same dimensions as the food cup in the larger G_1. The floor was black rubber matting. When *S* entered the smaller G_1 and approached the food cup at the front, a metal door was raised, restricting it to a $2 \times 2 \times 9$ in. space. The dimensions of the larger G_1 were $23 \times 3\frac{1}{4} \times 4$ in.

Design

The experiment was run in two stages. In the first (pretest) stage, *S*s were run to continuous reward in G_1 and to random presentation of the larger and smaller sizes of G_1. In the second (test) stage, *S*s were to have been split up into six independent groups (3×2 factorial) representing two degrees of confinement and three degrees of detention. In the light of pretest results (see Fig. 1) showing differential performance out of G_1 sizes within *S*s, it was necessary to continue running all *S*s to *both* sizes of G_1 throughout the test period also. Consequently, the design in the test period involves three independent groups for the detention variable, while for all *S*s the large vs. small G_1 variable remains a within-*S*s comparison. This amounts to a mixed design (Lindquist, 1953, p. 292).

Procedure

The entire experiment from the beginning of preliminary training involved a total of 130 days.

Habituation.—The preliminary training period was 22 days. Throughout this period *S*s were maintained on a 24-hr. food-deprivation schedule. Water was available at all times throughout the experiment. From Days 1 to 12 *S*s were fed 10 gm. Purina lab chow per day. From Days 13 to 18 all *S*s were handled for 3 min. prior to being fed. From Days 19 to 22, each *S* was placed in a carrying cage, carried to the experimental room, and placed in the apparatus for 5 min., during which time it was allowed to explore the apparatus with the photoelectric system operating, to become adapted to the clicking noises associated with the equipment.

Acquisition (pretest) training.—For a period of 80 days, each *S* was given one trial per day in the apparatus under approximately 23 hr. food deprivation. During acquisition *S*s were run to continuous reward and to random presentation of the smaller and larger areas of confinement in G_1. The order of G_1 sizes over 16 trials was as follows: LSSLLSLSSLSLSLLS, where L represents the large goal box and S the small. The *S*s were trained to both sizes of G_1 to reduce any aversiveness the smaller G_1 might have for them and to equalize their experience with both. The *S*s were run in squads of 12 and were fed in the home cage approximately 1 hr. following the daily trial.

The procedure on an individual trial was as follows: *S* was removed from its individual cage, placed in a compartment in a 12-place carrying cage, and taken to the experimental room where, for a period of 10 min. prior to running the first *S*, *E* checked the equipment and clocks, so that *S*s could become adjusted to the noises in the experimental room and so that stimuli preceding the first *S*'s trial would be the same as for any other *S*. Any acquisition trial was run as follows: (*a*) *S* was placed in SB and when it oriented toward the door, a black metal door was dropped. (*b*) Three seconds later a clear plastic door was lowered, allowing *S* to enter Runway 1. (*c*) The *S* traversed Runway 1 and entered G_1 which, depending on the trial, would be either large or small. On all trials, *S* received three 37-mg. Noyes food pellets in G_1. (*d*) After 15 sec., controlled automatically by an interval timer, a metal door was lowered, and when *S* had oriented toward it, a clear plastic door was lowered. (*e*) The *S* traversed Runway 2 into G_2 and received two 37-mg. food pellets. (*f*) Fifteen seconds after *S* entered G_2 it was removed and returned to the carrying cage.

Starting, running, and goal-entry measures were taken in both alleys of the apparatus, although for purposes of this report, starting and running measures in the first alley are not considered. The goal-entry time was examined to ascertain whether *S*'s entry into both sizes of G_1 had equalized. It was found that after about 40 trials, 20 exposures to each G_1 size, goal-entry performance into both sizes of G_1 remained about equal for the remainder of the acquisition phase of the experiment. All performance measures were taken by .01-sec. timers, activated by a photoelectric system. Performance measures in Runway 2 were taken as follows: Dropping the Plexiglas door leading into Runway 2 from G_1 activated a timer which stopped when *S* moved to a point 3 in. into Runway 2 and interrupted a photobeam, providing a starting time (latency) measure and starting a second timer. The interruption of a photobeam 25 in. from the start door stopped the second timer, providing a running-time measure, and started a third. A third photocell unit in G_2, 46 in. from the second unit, stopped the third timer providing the goal-entry measure.

Test trials.—Following the 80 acquisition trials, during which each *S* received 40 exposures to each of the two sizes of G_1, and was run to 100% reinforcement in G_1, a 24-trial test period was instituted. On these trials half the entrances into G_1 were rewarded and the other half nonrewarded to test for frustration effects. The *S*s were randomly assigned to one of three experimental groups—Groups 3, 15, and 90—which varied according to duration of detention, in seconds, on nonreward trials in G_1. "Detention" refers to the amount of time *S* spent in G_1 on nonreward trials of the test period. On the reward trials of the test period, all *S*s were detained for 15 sec. Throughout the test trials *S*s continued to run to both sizes of G_1, as in the earlier procedure.

In each block of eight test trials, each *S* had four trials to the large G_1, and four to the small. The reward and nonreward trials were arranged in such a manner that in any block of eight trials *S* had two reward and two nonreward trials in each of the two sizes of G_1. For instance, in a block of eight trials, if the order of goal box presentation was LSSLLSLS, the order of rewards and nonrewards was NRNRNRRN.

RESULTS AND DISCUSSION

Acquisition (pretest) training.—Starting, running, and goal times in Runway 2 over the 80-trial pretest phase are presented in Fig. 1. Each of the 10 points in each curve represents the mean of 56 individual *S* medians taken over blocks of 8 trials.

The curves in the left-hand panel show performance measures when *S* was entering Runway 2 following reward in the large G_1; whereas the curves in the right-hand panel show, *for the same* Ss, performance out of the small G_1 after reward. It is apparent that the starting- and running-time measures are grossly affected by size of G_1 out of which *S* runs, while the goal-time measure seems entirely unaffected by size of G_1. It is clear that, *within the same* Ss, the starting and running performance measures depend on size of G_1 out of which *S* starts into Runway 2. These effects seem to persist undiminished over the 80 trials of this phase of the experiment. The group data presented in Fig. 2 are characteristic of almost every *S* in the experiment.

Test trials.—In the course of 24 test trials, each *S* had 6 reward and 6 nonreward trials in the small G_1, as well as 6 reward and 6 nonreward trials in the large G_1. A median of the 6 trials under each of these four conditions was computed for each *S*, and the points on the graph represent the mean of such 6-trial medians for all *S*s in each detention group. Each panel of Fig. 2 shows a comparison of Runway 2 performance in the last block of preliminary training, and on R and N test trials for the three detention durations, under one of two G_1 sizes, and for one of three performance measures. Within each panel between-groups comparisons are involved; however, comparisons across panels, left and right, are always within *S*s.

The two top panels show that starting-time measures, in general, were higher under large G_1 conditions than under small G_1 conditions. The middle panels show the opposite for running time. These findings are in line with the pretest phase of the experiment. The FE is most pronounced for starting measures out of the large G_1 and for running measures out of the small G_1. These two are, then, the measures of importance in regard to evaluating the detention variable. The FEs for goal time are very small in five cases, and negative

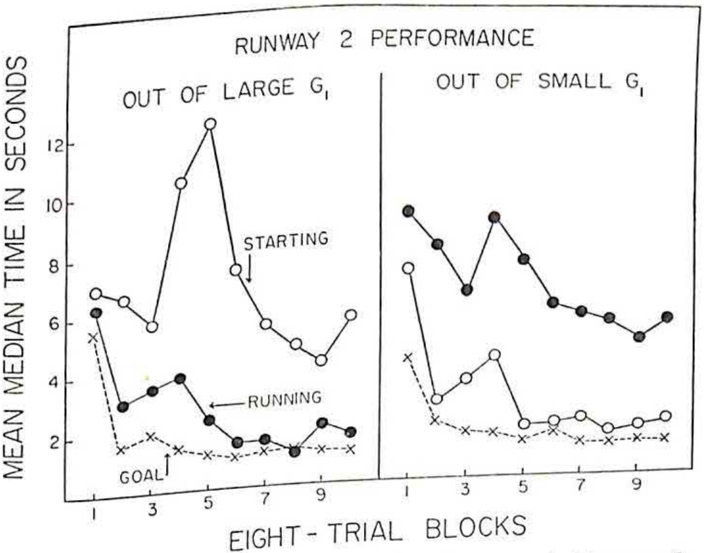


FIG. 1. Starting, running, and goal performance in Runway 2 when *S*s ran out of the larger or the smaller G_1.

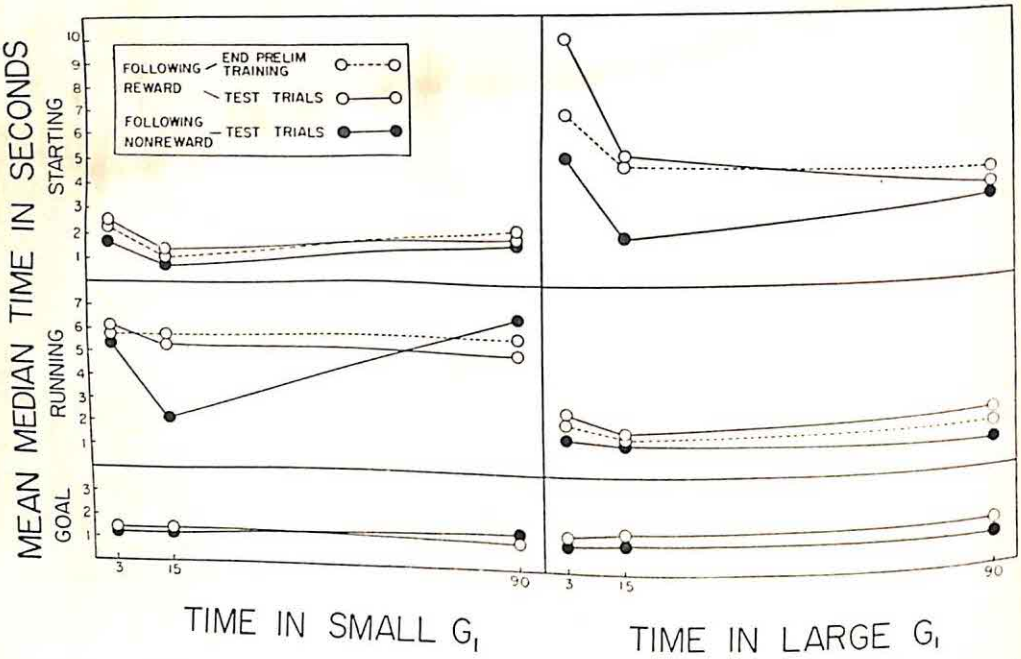


FIG. 2. Test-trial performance in Runway 2 as a function of time in larger and smaller G_1 on nonreward test trials. (Comparison reward curves of two kinds—open circles—are also shown.)

in the other. Under the large G_1 condition, the greatest discrepancy (FE) between reward and nonreward test performance appears to be in Group 3. This is due to an apparent relationship between R and N test-trial performance which seems particularly evident in the starting measures. The large discrepancy between R performance on test trials and on terminal pretest trials in the Group 3 starting measure out of the large G_1 is not evident in any of the other detention groups, for any measure, in either G_1 condition. The relationship of detention to size of FE in the other measure of interest—running out of small G_1—shows the 15-sec. condition clearly different from the other two.

A mixed-design analysis of variance was performed separately on the starting and running measures. For start time, the interaction between G_1 size and reward-nonreward was significant at the .05 level, $F\ (1, 53) = 6.00$. The significant interaction would suggest that the FE varies with size of G_1. No other interaction was statistically significant. The difference between starting-time measures under the large and small G_1 conditions was significant, $F\ (1, 53) = 23.70$, $p < .01$, as was the reward vs. nonreward (frustration) comparison, $F\ (1, 53) = 13.13$, $p < .01$.

The between-*S*s comparison showed that starting time varies with time spent in G_1 on the nonreward trials, $F\ (1, 53) = 3.74$, $p < .05$. Time of detention in G_1 did not interact significantly with the reward-nonreward variable, for the starting measure.

Wilcoxon matched-pairs signed-rank tests were used to evaluate individual reward-nonreward differences in the starting-time measure. For large G_1 comparisons, Group 3 ($T = 30$, $p < .01$) and Group 15 ($T = 10$, $p < .01$) yielded significant differences, and the difference for

Group 90 is not significant ($T = 58$). For small G_1 comparisons, Group 3 ($T = 28$, $p < .01$) and Group 15 ($T = 22$, $p < .02$) yielded significant differences in the direction of the FE. Again, the difference for Group 90 was not significant ($T = 75$).

The analysis of variance of the running-time data resulted in a significant triple interaction among Reward, Confinement, and Detention conditions, $F(1, 53) = 6.79, p < .05$. As in the preceding analysis, reward conditions, $F(1, 53) = 5.43$, $p < .05$, and size of G_1, $F(1, 53) = 32.84, p < .01$, were significant, confirming the relevancy of reward vs. nonreward and size of G_1 as variables affecting Runway 2 performance.

Wilcoxon matched-pairs signed-rank tests were again used to test for significance of the FE in individual detention groups. In each detention group, under the large G_1 condition, the differences were statistically significant in the direction of the FE ($T = 25, p < .01$ for Group 3; $T = 21$, $p < .01$ for Group 15; and $T = 45$, $p < .05$ for Group 90). For the small G_1 condition, the FE was significant only for Group 15 ($T = 29, p < .05$).

In the case of the third measure, goal time (bottom panels of Fig. 2), the graphs indicate only slight differences between performance on reward and nonreward test trials. For this measure, performances for the three detention times are very similar, and the effect exerted by the confinement variable on the other two measures appears to dissipate as *S* traverses the runway, particularly out of the small G_1.

This experiment demonstrates the FE with a one-trial-per-day procedure. The results also demonstrate that the absolute size of the FE is greater for those measures which represent relatively weaker performance, i.e., the starting-time measure out of the large G_1, and the running-time measure out of the small G_1. This is reasonable in terms of a "ceiling" effect.

The hypothesis offered by Amsel and Roussel (1952), that the facilitating effect of nonreward on subsequent performance should vary with time in the frustrating situation, has received some support from this study, at least to the extent that the FE seems greatly attenuated by 90 sec. The more specific suggestion emerging out of their study, that the greatest FE should appear following some intermediate time of detention in G_1 on nonreward trials, is not strongly supported by the present results. While the running-time data out of small G_1 would, by themselves, tend to support such a conclusion, the starting data out of large G_1 do not. Here, the test-trial reward value at 3 sec. detention is troublesome, suggesting the possibility of a carry-over of the detention effect, at this value of detention anyway, from nonreward to reward trials.

A very stable finding in this experiment was that, in the double-runway situation, size of G_1 is an important determiner of absolute level of Runway 2 performance, at least with the extremes of size used here. The pretest curves clearly show that starting times are shorter out of the small G_1 than out of the larger G_1 for the same *S*s; and that running times are longer out of the small G_1 than out of the larger G_1. All trials during acquisition were reward trials and *S*s were detained for 15 sec. in both sizes of G_1. It should be noted that, after the first few training trials, all *S*s were able to consume the three 37-mg. pellets in G_1 in less than 15 sec. Consequently, on most trials, *S*s were required to wait for a few seconds for the lowering of the plastic door into Runway 2. This suggests that the slower starting times out of the larger G_1 may simply reflect a greater opportunity for competing responses in the larger G_1. An alternative, although not incompatible, hypothesis is that the very small G_1 was so restricting as to be aversive, yielding fast starting (escape) times.

Running was slower following small G_1 confinement, just as starting was faster. Once *S*s left the small G_1, they frequently stopped just after breaking the photoelectric beam which started the running-time clock. Consequently, the running measure out of the small G_1 often involved a "restarting time." This may simply reflect an increased tendency to explore following confinement. These same *S*s seldom paused in Runway 2 after leaving the large G_1.

The results of the present experiment indicate much weaker overall performance (i.e., slower starting and running speeds) than in previous studies employing the same double-runway apparatus. It is suggested that, apart from the obvious effect of varying G_1 size, the poorer performance may be due to the fact that *S*s were run one trial a day to relatively small rewards. (In experiments involving more than one trial a day, there is almost always a "warm-up" effect, the first-trial performance being somewhat weaker than that on subsequent trials.) However, the weaker performance levels also mean larger absolute FEs, which suggests that for parametric experiments, in which size of FE is the dependent variable, the one-trial-per-day procedure may be a good one to employ.

REFERENCES

AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychol. Rev.*, 1962, **49**, 306–328.

AMSEL, A., & PENICK, E. The influence of early experience in the frustration effect. *J. exp. Psychol.*, 1962, **63**, 167–176.

AMSEL, A., & ROUSSEL, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

(Received June 19, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 475–482

ABSOLUTE JUDGMENTS OF DISCRETE QUANTITIES RANDOMLY DISTRIBUTED OVER TIME[1]

DWIGHT E. ERLICK

Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio

The accuracy with which S can make estimates of the percentage of sequential events falling in 1 of 2 categories as a function of: (a) the true percentage being estimated and (b) the length of the sequence was investigated. The highly discriminable events were presented in a random sequence at the rate of 4/sec. 2 groups of 100 Ss each and 4 individual Ss furnished the data. Results indicate: (a) decrease in error of estimation as length of sequence increases, (b) systematic overestimation of percentages below 50% and underestimation of percentages above 50%, (c) higher accuracy of estimation for extreme percentages (high-low) than for the intermediate percentages, (d) mean accuracy of 77% correct estimations when 4 response categories are utilized with a systematic decrease in accuracy as the number of response categories increases, (e) response bias with percentages of 40, 50, and 60 being used most frequently.

Earlier investigations have been concerned with several variables related to the a posteriori perception of *relative* frequency of random sequential events (Erlick, 1960, 1961, 1962). In these studies, highly discriminable events were presented randomly at fast rates (4/sec) and *S* was required to determine which event had occurred most frequently. The findings indicated a linear relationship existed between the information transmitted and the frequency difference between events; also there was a general overall increase in accuracy of perceiving the more frequent event as the observation time increased from 5 to 40 sec.

Shuford (1961) using a spatial display had *S*s estimate the percentage of one of two distinct elements viewed in a 20 × 20 matrix. Two durations of observation were used: 1 sec. and 10 sec. The results indicated an increase in the variability of the estimates for the shorter duration of observation. However, accuracy of estimations did not appear to be affected by the duration of observation. Also, the ability to make estimates did not appear to be systematically affected as a function of the actual percentage being estimated and in general was quite accurate throughout the entire range of percentages.

The present investigation was concerned with the absolute estimation of the frequency of occurrence of random sequential events: three factors were varied: the percentage being estimated; the length of sequence; and the variability of the length of sequence.

METHOD

A punched paper tape reader was used to present a preprogramed series of the two letters A and C which were displayed on the same location on a 3 × 4 in. screen by means of a rear projection system. The letters, $3\frac{3}{4}$ in. high, were presented one at a time at the rate of four per second and were on for approximately half of each cycle, producing an equal on-off ratio. This rate and ratio prevented *S* from counting the individual letters, yet produced a distinct off-time when the same symbol was repeated in sequence.

The total series of the letters A and C presented within one trial was one of five lengths, i.e., 20, 40, 60, 120, or 240. When presented at the rate of four per second, these resulted in total observation times per trial of 5, 10, 15, 30, and 60 sec., respectively, and will be referred to as T_5, T_{10}, T_{15}, T_{30}, and T_{60}, respectively. Within each T condition 21 different ratios between the letters A and C were presented. Nineteen of these were selected so as to increase the percentage of occurrence of one of the letters in 5% increments from 5 to 95%. Thus in Cond. T_5, 4 A's and 16 Cs present 20% A's and 80% Cs. Since the percentages of the two events A + C = 100%, each percentage condition was counterbalanced in terms of the symbols (A and C) by varying the instructions as to which symbol was to be estimated.

---

[1] The data were collected at the Engineering Psychology Research Project, Antioch College.

To evaluate estimates of very large percentage differences, two ratios were presented within each T condition (except T_5) in which all letters were the same except one, e.g., in T_{15} 1 A and 59 Cs or in T_{60} 1 A and 239 Cs. These latter two conditions resulted in percentage of occurrences ranging from 0.4% to 2.5% or 97.5% to 99.6%, depending on the T condition and which letter was estimated. A 0% and 100% condition was presented under the T_5 condition. The five T conditions and the 21 ratios within each resulted in 105 experimental conditions.

A total of 200 *S*s was tested. For 100 *S*s all 105 conditions were presented in a random order, i.e., the duration (T condition) was variable from trial to trial (V group). For the other 100 *S*s each of the five T conditions was presented separately in a counterbalanced Latin-square design to five groups of 20 *S*s each, with the 21 ratios within each T condition occurring randomly, i.e., the durations were constant from trial to trial for a series of 21 ratios (C group). A trial required one judgment and consisted of observing a random series of one of the ratios between the letters A and C.

The *S*s were told to estimate the percentage of the specified letter to the best of their ability. They were told to guess if they were not sure of the answer and to keep all previous answers covered with a card with which they were provided. All 105 trials were punched on one continuous tape with an automatic stop programed between trials. The *E*, who sat in the room with *S*s, called out the number of each trial before it was presented and started the trial by pressing a button. Each of the 200 *S*s made 105 percentage estimates resulting in an overall total of 21,000 responses. The *S*s were run in groups of up to 10 at one time, with each *S* seated in an individual booth facing the display and viewing the letters from a distance of from 6 to 9 ft. The *S*s were given two practice trials before each session. An attempt was made to motivate *S*s by offering a monetary bonus as a function of the accuracy of their estimations.

To determine if a few observations averaged over many *S*s produced some of the same functional relationships as many observations made by one *S*, four *S*s from the group experiment were run independently of the foregoing. The length of sequence for all the conditions was the same, 60 (T_{15}). Each of the four *S*s took a series of 10 experimental sessions consisting of 10 different random replications of the 21 different per cents to be estimated, thus giving a total of 2,100 observations, i.e., 100 observations for each of the 21 different per cents. The selection of the four *S*s was based on their better than average performance in the group task, as indicated by their overall absolute error based on the difference between the actual and estimated percentages.

RESULTS

The following analysis of the two groups (C and V) was based on the overall mean of the 100 *S*s within each group. Likewise, the analysis of the individual *S*s was based on the overall mean of all four *S*s.

Reversal errors.—For some trials it appears that *S* might have temporarily forgotten which symbol he was estimating because his estimates would be approximately the equivalent of the alternate symbol he was not supposed to be estimating. This was most apparent for the percentages at the extremes of the scale where 5% might be estimated somewhere between 90% and 95% and 90% might be estimated somewhere between 5% and 10%. For percentages presented near the middle of the scale (40–60) it was difficult to differentiate perceptual errors from reversal errors. In order to reflect mainly the effects of perceptual errors and since an error of 40 was several standard deviations from the mean response in most instances, it was decided to arbitrarily call all errors of 40 or greater, reversal errors. The overall numbers of reversal errors for the V and C groups were approximately the same 3.1% and 2.9%, respectively, while the four individual *S*s had an overall average of 2.1%. In all the following analyses, reversal errors have been omitted. While this tends to disproportionately reduce the mean error and variability for extreme percentage estimations, the small number of reversals did not affect any of the general functions reported.

Length of sequence.—There was no systematic interaction between the effects of length of sequence (duration of observation) and the magnitude of the percentage observed. The mean absolute percentage error decreased for both C and V groups as the length of sequence increased. The initial percentage error at 5 sec. of observation was higher for V (10%) than C (8%) and systematically decreased leveling off for both groups at approximately 7% for 30 sec. of observation. The same data when interpreted in terms of information transmitted results in an increase in the transmission of information as a function of increasing the length of sequence with the curve asymptoting at slightly less than 2 bits at 30 sec. of observation, i.e., 120 events (see Fig. 1).

Practice effects.—A comparison of the first five experimental sessions with the last five for each of the four individual *S*s showed no systematic differences in error as a function of the 21 different percentages presented for estimation.

Magnitude of percentage observed.—For both groups and individual *S*s there was a systematic overestimation of percentages below 50% and a systematic underestimation for percentages above 50% (see Fig. 2). The mean estimations for 50% were quite accurate.

The σ for the four individual *S*s, derived by determining the σ for each *S* and averaging the four σ's, was

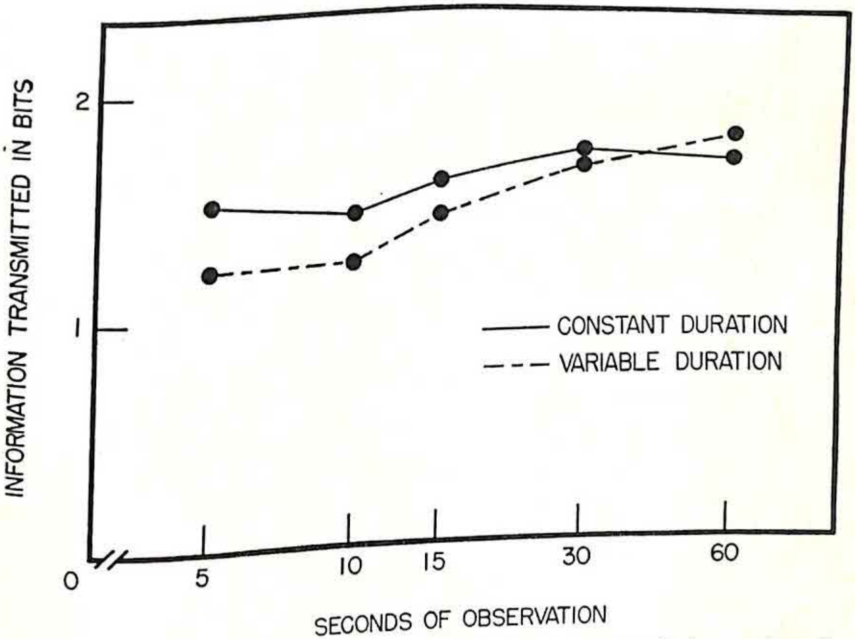


FIG. 1. Information transmitted as a function of duration of observation (length of sequence) for conditions when the trial-to-trial durations were constant and variable. (The abscissa is scaled logarithmically. Each point is based on 2,100 observations.)

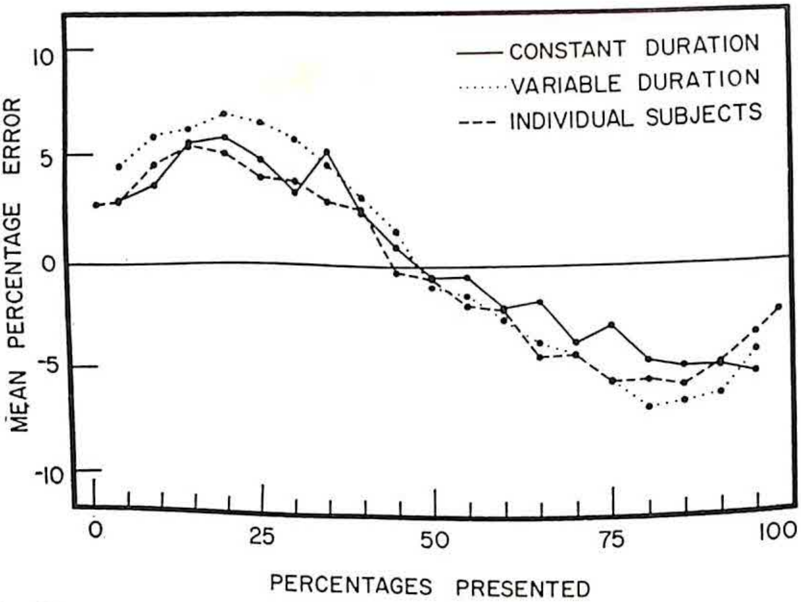


FIG. 2. Mean percentage error as a function of percentages presented for constant- and variable-duration groups and individual *S*s. (Each point on the group curves is based on 500 observations, while each point on the individual curve is based on 400 observations.)

lower for those percentages at the extremes of the scale, i.e., 0–20 and 80–100. The σ for the extreme ends of the scale of 5% and 95% was approximately 5, this systematically increased with the variability being approximately 9 over the middle range from 20% to 80%.

Correct identifications.—The individual *S*'s responses were placed in 21

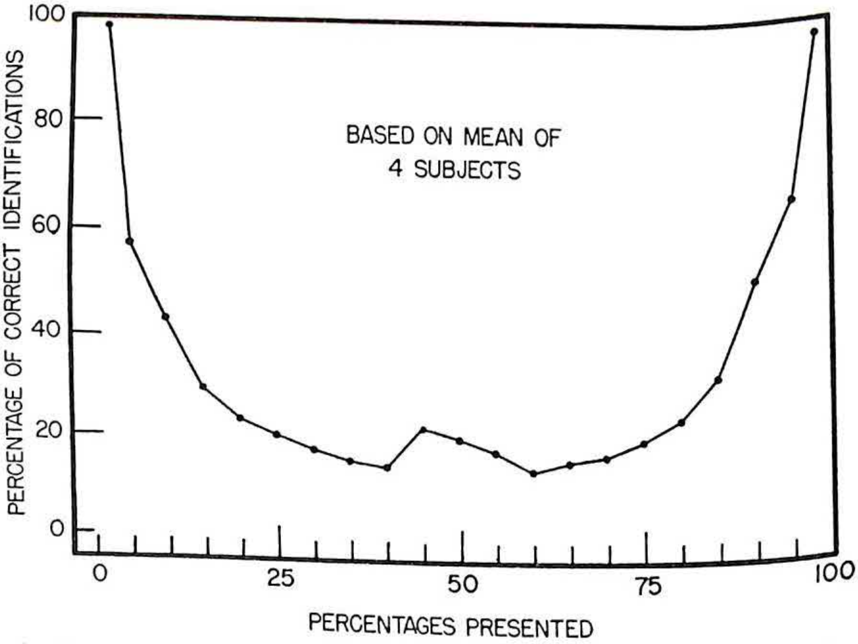


FIG. 3. Percentage of correct identifications as a function of the percentages presented. (This measure was computed by dividing the number of times a response was used correctly by the total number of times the response occurred.)

categories corresponding to the 21 stimulus categories presented. Responses occurring within the corresponding stimulus category were considered correct identifications. An analysis of the percentage of correct identifications based on the total number of responses made within each category was made. Higher accuracy of estimation was found for the extreme percentages (high-low) than for the intermediate percentages (see Fig. 3).

Information transmitted.—The information transmitted by each of the individual *S*s ranged from 1.55 to 2.20 bits with an overall average of 1.94 bits.

Number of response categories.—The stimulus scale was transformed into an equal-discriminability scale (EDS) for each of the four *S*s (Garner & Hake, 1951). To determine the effects of varying the number of response categories, the EDS was then divided into *N* (2, 3, 4, 5, 7, 10, or 21) equal-discrimination categories. Using response intervals corresponding to the stimulus intervals, the percentage of correct identifications based on the total number of responses made within the specified interval was determined for each of the intervals within each *N* condition. The mean percentage for all intervals within each of the *N* conditions was then determined. This mean percentage of correct responses decreased as the number of response categories increased. Approximately the same degree of accuracy was obtained with equal intervals along the stimulus scale as with the EDS procedure, therefore the equal-interval method of categorizing was used since it was more parsimonious (see Fig. 4).

Bias correction.—It is apparent from Fig. 2 that both groups and individual *S*s show the same systematic bias. Thus, an analysis was done to determine the predictability of individual *S* responses using the error of

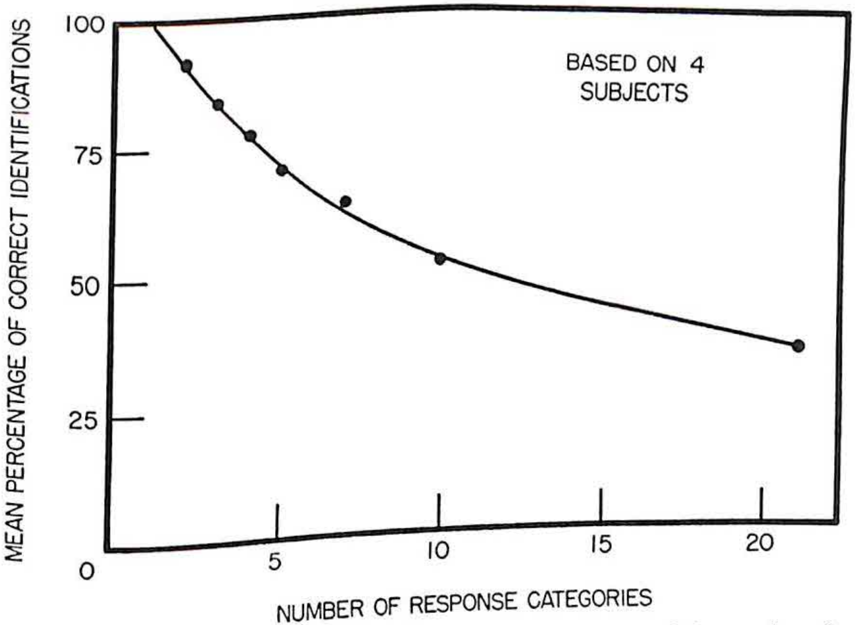


FIG. 4. Mean percentage of correct identifications as a function of the number of response categories. (Response categories were imposed upon the data and were selected in terms of equidistant units along the stimulus scale.)

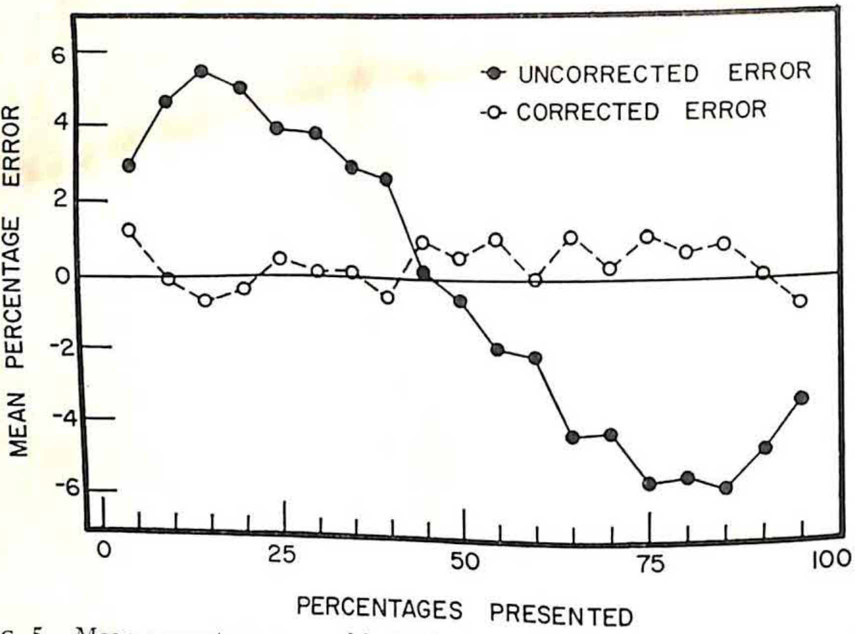


FIG. 5. Mean percentage error of individual *S*s as a function of percentages presented. (Uncorrected error is based on difference between percentage presented and the mean of the four individual *S*s' responses. Corrected error is based on the difference between the mean of the four individual *S*s' responses and a prediction of these responses based on the constant-duration group.)

the constant-duration group as a correction factor. Since the degree of overestimations below 50% and the underestimations above 50% is approximately the same, the mean error for stimuli equidistant on both sides

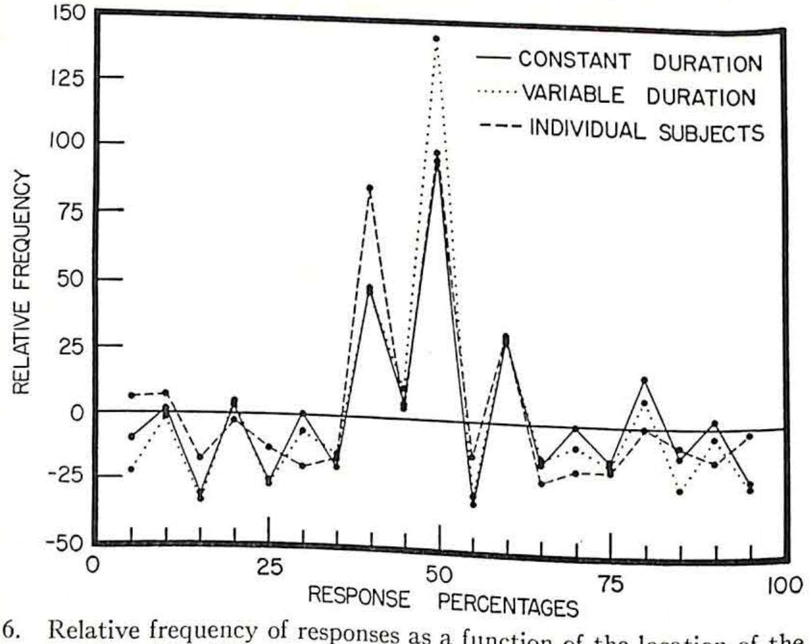


FIG. 6. Relative frequency of responses as a function of the location of the response on the percentage scale. (The "0" assumes an equal distribution of responses for all 21 response categories. Relative frequency is the percentage increment over the expected "0" value. The "0" value for the groups is 500 and for the individual *S*s, 400.)

of 50% was averaged. A quadratic equation relating these averaged errors as a function of the distance of the stimulus from 50% was derived from the mean error of the constant-duration group. The following equation indicates the correction (C) to be added or subtracted from a stimulus percentage ($S\%$) over the range of 5–95%, to predict the response percentage that an observer will make.

$$C = -.0048X^2 + .3247X - .4504$$
$$X = \%\text{ deviation of } S\% \text{ from } 50$$
$$C \begin{cases} \text{added when } S\% < 50 \\ \text{subtracted when } S\% > 50 \end{cases}$$

Using the foregoing equation, there is a very significant reduction in the absolute mean error of the individual *S*s from 3.63% to .62%, $F(19, 19) = 5.57$, $p < .001$. (See Fig. 5.)

Response bias.—An analysis of *S*s' responses was made to determine the extent of any preferences or response biases. Assuming an equal distribution of responses for all 21 response categories, the percentage of this expected value (400 for four *S*s and 500 per group) was determined for each of the 21 percentages presented. It is apparent from Fig. 6 that *S*s have a marked tendency to use the middle of the scale more frequently than the extremes, especially responses of 40, 50, and 60. This bias is true for both groups as well as the individual *S*s.

DISCUSSION

The overall percentage of reversal errors agrees with previous findings of 2.1% for reversal errors in the reading of air navigation plotters (Christensen, 1946). This tendency to use the "obviously inappropriate" response might possibly be due to inattention.

The general increased information transmitted (I_t) as a function of increased observation time confirms similar findings for the perception of *relative* frequency of two discrete events. Also, the difference in the constant and variable duration groups in terms of I_t at the shorter sequences is similar to that found in an earlier study by the author using measures of relative frequency (Erlick, 1960, 1961). The lack of a practice effect is not too surprising since *S*s were not given any knowledge of results.

The fact that both groups and individual *S*s overestimated percentages below 50% and underestimated percentages above 50% is not unusual, since this might be expected to some degree at both extremes of a finite scale. However, the highly consistent similarity of the three curves (Fig. 2) is rather striking. While the *mean* error is greater for the extreme high and low percentages than for the middle percentages, this is somewhat misleading as can be seen in terms of the percentage of correct identifications in Fig. 3. Similar findings have been found in other studies of absolute judgments (Eriksen & Hake, 1956). What happens is that a greater number of responses are made in the center portions of the scale (Fig. 6). These responses tend to be equally distributed about the mean, making the mean response highly accurate, the variability quite high, but the percentage correct rather small.

The information transmitted was approximately 2 bits. While in theory this would indicate the ability to absolutely identify four categories of frequency difference, it is apparent from Fig. 4 that this is not the case. This curve based on the tabulation of "correct responses" indicates about 77% mean accuracy of identification for four categories. The discrepancy is probably due to the large variability of responses.

The highly significant reduction in the mean error of the individual *S*s suggests the feasibility of utilizing the correction equation for correcting percentage estimates, at least for untrained *S*s. Also, it suggests the possibility of utilizing this correction factor in terms of training naive *S*s.

Responding more toward the middle of a finite scale, suggests a tendency to favor the mean when in doubt. The fact that percentages of 40, 50, and 60 were highly preferred also confirms previous findings of favoring a "0" in the final digit of unstructured numerical responses (Cohen & Senders, 1954).

REFERENCES

CHRISTENSEN, J. M. Psychological factors involved in the design of air navigation plotters. *USAF AMC memo Rep.*, 1946, No. TSEAA-694-1D.

COHEN, J., & SENDERS, V. L. Factors affecting the frequency of various final digits. *USAF WADC tech. Rep.*, 1954, No. 55-371.

ERIKSEN, C. W., & HAKE, H. W. Anchor effects in absolute judgments. *USAF WADC tech. Rep.*, 1956, No. 56-144.

ERLICK, D. E. Judgments of the relative frequency of two random sequential events: Effects of duration of observation. *USAF WADC tech Rep.*, 1960, No. 60-673.

ERLICK, D. E. Judgments of the relative frequency of a sequential series of two events. *J. exp. Psychol.*, 1961, **62**, 105–112.

ERLICK, D. E. Perception of the most frequent category of a random series as a function of the number of categories. *J. exp. Psychol.*, 1962, **63**, 115–118.

GARNER, W. R., & HAKE, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.

SHUFORD, E. H. Percentage estimation of proportion as a function of element type, exposure time, and task. *J. exp. Psychol.*, 1961, **61**, 430–436.

(Received June 24, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 483–485

DIRECT COMPARISON OF AUDITORY AND VISUAL DURATIONS[1]

SANFORD GOLDSTONE AND JOYCE LEVIS GOLDFARB

Baylor University College of Medicine

It has been shown repeatedly that longer visual than auditory durations are judged equivalent to temporal concepts such as a clock second. The present experiment explored this intersensory difference in time judgment through cross-modal and intramodal successive comparisons of auditory and visual inputs. Auditory durations were judged longer than visual, and visual durations were judged shorter than auditory for 3 different interstimulus intervals. The intersensory phenomenon previously obtained with the absolute method was also found by direct comparison suggesting a fundamental cross-modal difference characteristic of human time judgment.

When *S*s made single-stimuli estimates of lights and sounds as either *more* or *less* than their concept of one clock second, longer visual than auditory durations were judged equivalent to the temporal concept; visual inputs were judged shorter than auditory inputs of the same physical duration (Goldstone, Boardman, & Lhamon, 1959). Experiments have demonstrated that this curious intersensory difference in time judgment is independent of psychophysical method, stimulus characteristics such as pitch, hue, and intensity, size of and distance from visual target, position in a series and number of response categories; it is present under various background anchor conditions, within very narrow ranges of short durations and in broader ranges up to 5 sec., and with different internal temporal standards (e.g., Behar & Bevan, 1961; Goldstone & Goldfarb, 1964).

Although this auditory-visual difference in time judgment has been demonstrated repeatedly with the method of absolute judgment, no experiment has investigated whether auditory durations are judged longer than visual, and visual shorter than auditory upon direct comparison. It is possible that this intersensory phenomenon is a characteristic of the method of single stimuli which requires *S* to compare a light or sound with an internal temporal standard, and not a fundamental cross-modal difference in human time judgment.

In this experiment *S*s compared the duration of variable lights with a standard sound, variable sounds with a standard light, variable sounds with a standard sound, and variable lights with a standard light within the same experimental context and using three standard-variable interstimulus intervals. It was possible to determine whether sounds were judged longer than lights, and lights shorter than sounds in relation to intramodal control conditions. The effect of increased interstimulus interval upon the intersensory difference was also explored.

METHOD

An electronic timing circuit controlled the duration of the lights and sounds and programed a constant interstimulus interval. An audio oscillator and headphones provided tones of 1,000 cps at 94 db. (re .0002 dyne/cm^2). A fluorescent system illuminated a 1-in. opal glass circular target mounted 43 in. from *S*'s eyes with a blue white light of 5 apparent

---

[1] Supported by United States Public Health Service Grant MH 01121.

ft-c. The *S* was seated in an acoustic chamber that reduced ambient stimulation.

The *S*s received pairs of durations and compared the second with the first using a 9-category scale from 1, *very much shorter*, to 9, *very much longer*, with 5, *equal*, in the middle. The first duration or standard was always 1 sec. and the variable was either .6, .8, 1, 1.2, or 1.4 sec. The five standard-variable combinations were randomly presented 10 times for each of the following four sense-mode orders: Audition, standard-audition variable; vision, standard-vision variable; audition, standard-vision variable; vision, standard-audition variable. The *S*s received 200 pairs with sense-mode order randomized within the series and all pairs were separated by a 5-sec. interval.

Sixty college students were divided into three equal groups with a standard-variable interstimulus interval of either 1, 2, or 4 sec. Each *S*'s average category response for each standard-variable combination (V) and each sense-mode order (M) was obtained for the three interstimulus intervals (I).

TABLE 1

ANALYSIS OF VARIANCE OF COMPARATIVE JUDGMENTS FOR THE SENSE-MODE ORDERS (M), VARIABLE DURATIONS (V), AND INTERSTIMULUS INTERVALS (I)

Source	*df*	*MS*	*F*
Interval (I)	2	2.37	1.60
Error (b)	57	1.48	
Variable (V)	4	474.25	1030.97***
Mode (M)	3	30.98	67.34***
V × M	12	1.00	12.50***
I × V	8	.95	2.07*
I × M	6	.63	1.36
I × V × M	24	.11	1.37
Error$_1$ (w) V/I	228	.46	
Error$_2$ (w) M/I	171	.46	
Error$_3$ (w) VM/I	684	.08	

* $p < .05$.
*** $p < .001$.

RESULTS

Figure 1 displays the mean responses of the three I groups for each V and M condition. The curves and the analysis of variance summarized in Table 1 show that the auditory durations were judged longer than

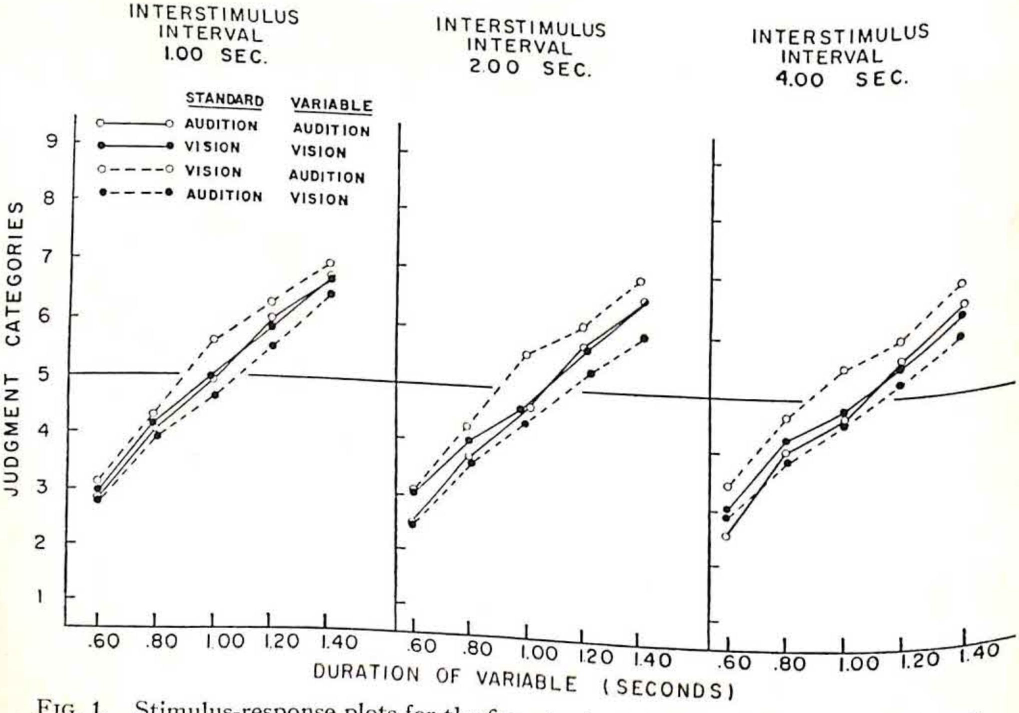


FIG. 1. Stimulus-response plots for the five standard-variable (V) combinations and the four sense-mode orders (M) of the three interstimulus-interval (I) groups.

visual, and the visual durations were judged shorter than auditory for the cross-modal sensory orders of the three groups. The intramodal curves were between those of the cross-modal conditions, and exhibited a positive time-order error. Thus the previously reported auditory-visual difference in time judgment obtained using the absolute method was demonstrated with direct comparison. A steeper slope for cross-modal and intramodal auditory judgments was reflected in the M × V interaction, and the slope flattened with the 4-sec. interstimulus interval producing the I × V interaction. The auditory cross-modal curves differed more from the intramodal curves than did the visual for all groups, and increased interstimulus interval did not modify the intersensory difference.

This experiment adds to the accumulating evidence that there is a fundamental intersensory difference characteristic of human time judgment; auditory durations are judged longer than visual, and visual durations are judged shorter than auditory.

REFERENCES

Behar, I., & Bevan, W. The perceived duration of auditory and visual intervals: Cross-modal comparison and interaction. *Amer. J. Psychol.*, 1961, **74**, 17–26.

Goldstone, S., Boardman, W. K., & Lhamon, W. T. Intersensory comparisons of temporal judgments. *J. exp. Psychol.*, 1959, **57**, 243–248.

Goldstone, S., & Goldfarb, J. L. Auditory and visual time judgment. *J. gen. Psychol.*, 1964, **70**, 369–387.

(Received June 24, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 486–494

EFFECTS OF NONLINEAR AND DISCRETE TRANSFORMATIONS OF FEEDBACK INFORMATION ON HUMAN TRACKING PERFORMANCE[1]

DARWIN P. HUNT

University of Dayton

9 groups of 7 male college students each were given 40, 1-min. trials on a complex compensatory tracking task under 9 different informational display conditions: 3 levels of informational quantization (3, 7, or 11 informational categories) were combined with 3 types of overall relationships between the displayed error and the actual error (nonlinear minification, linear, or nonlinear magnification). An index of tracking accuracy and of tracking economy was obtained by integrating the absolute values of voltage analogues of the actual tracking error and the control-stick deflections, respectively. Greater tracking accuracy was associated with (a) the linear informational display, (b) a larger number of informational categories, and (c) practice. Practice differentially influenced tracking economy depending upon the number of informational categories. The amount of practice given resulted in a relatively stable terminal level of performance for tracking accuracy, but not for tracking economy.

Several previous investigations have been concerned with the effects of either nonlinear or discrete transformations of feedback information on the human performance of psychomotor tasks. Bilodeau (1953) has studied the influence of variations in the slope of linear transformations of feedback information on the performance of the relatively simple task of making blind settings on a micrometer. Noble and Broussard (1955) extended Bilodeau's work to include curvilinear transformation. Schoeffler (1955) related performance on a simple tracking task to the level of informational quantization. In a previous study (Hunt, 1961) the effects of discrete transformations of error information on the performance of a complex compensatory tracking task were investigated.

In all of the studies indicated above, the independent variable has involved a variation in the linear magnification *or* the curvilinear nature of the informational transformation *or* the quantization of overall linear transformations. The present research addresses itself to the particular question, "What are the effects of curvilinear *and* discrete transformations of information on human compensatory tracking performance?"

METHOD

Subjects.—Sixty-three right-handed male college students who had no prior experimental tracking experience were paid to serve as *S*s.

Apparatus.—The apparatus consisted of a forcing-function generator, a lever control, an error transformer, an oscilloscope, and an electronic computer (Fig. 1). The forcing-function generator provided a voltage which, if not compensated for by appropriate move-

---

[1] The data were collected under United States Air Force Contract No. AF-33(616)-7863 while the writer was employed at the Behavioral Sciences Laboratory, Wright-Patterson Air Force Base. The valuable comments of M. J. Warrick are gratefully acknowledged. James Nehez, University of Dayton, constructed and maintained the apparatus and some of the data analysis was carried out at the Behavior Research Laboratory, Antioch College under Contract No. AF 33(657)-7362.

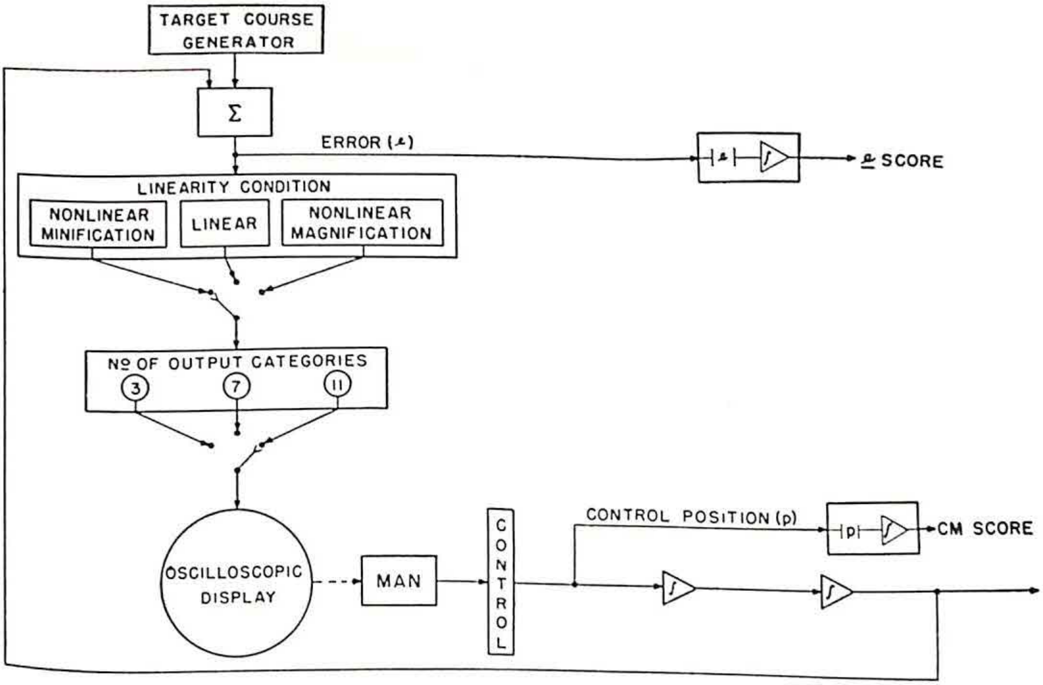


FIG. 1. Block diagram of the apparatus.

ments of the control stick by *S*, would drive a dot right and left on an oscilloscopic display. This forcing function was composed of three sine functions: 1 cpm, 2 cpm, 3 cpm with the 3-cpm component 180° out of phase.

The *S*'s control was a 23.5-in. lever with its pivot point 3.5 in. above the floor directly in front of him. When the lever was in a center upright position, its tip was 10 in. in front of and 18 in. below the center of *S*'s display. The control could be moved laterally 20° to the right or left, but was fixed in the fore-aft direction. When no sideward force was applied to the control, slight spring centering held the lever in the neutral position at which the control output was zero. The voltages resulting from displacements of the lever from the upright position were linearly related to the angular deflection of the lever.

The *S*'s display consisted of a 1 in. high × 4.5 in. wide exposed portion of the face of the oscilloscope. A single vertical illuminated line, centered horizontally in the 4.5-in. slot, identified the zero error position for *S*. The transformed error was encoded and displayed within the slot as a small dot which moved in a discontinuous fashion along the horizontal dimension.

The computer served two general purposes: To provide the controller dynamics and to obtain performance scores. The controller dynamics were obtained by a double integration of the voltages produced by deflection of the control lever; thus, *S* was operating an acceleration control. The integration rates of the two integrating amplifiers were such that a 3° deflection of the control lever alone, i.e., ignoring the forcing function, would result in an error increasing .32 in/sec^2, e.g., in 3 sec. the error would be 1.44 in.

The performance of *S* was measured in two ways: (*a*) integrated control motion (CM) and (*b*) integrated error (e). The CM scores were obtained by feeding the absolute value of the voltage analogue of the control lever deflections, i.e., stick-deflection voltages, into the computer, which integrated these values for each trial (see Fig. 1). The e scores were obtained in the same way, except that voltage analogues of the true error were used instead of the stick-deflection voltages. These e and CM scores were available to *E* on digital displays.

The instantaneous difference between the voltage produced by the forcing-function generator and a double-integrated voltage resulting from displacements of the control stick is called the true error. This true error was fed into the error transformer and was transformed according to the informational condition under which the particular *S* was scheduled to be tested. Nine transformations of the true-error information were employed (Fig. 2). In all cases, a true error of 0 ± 0.013

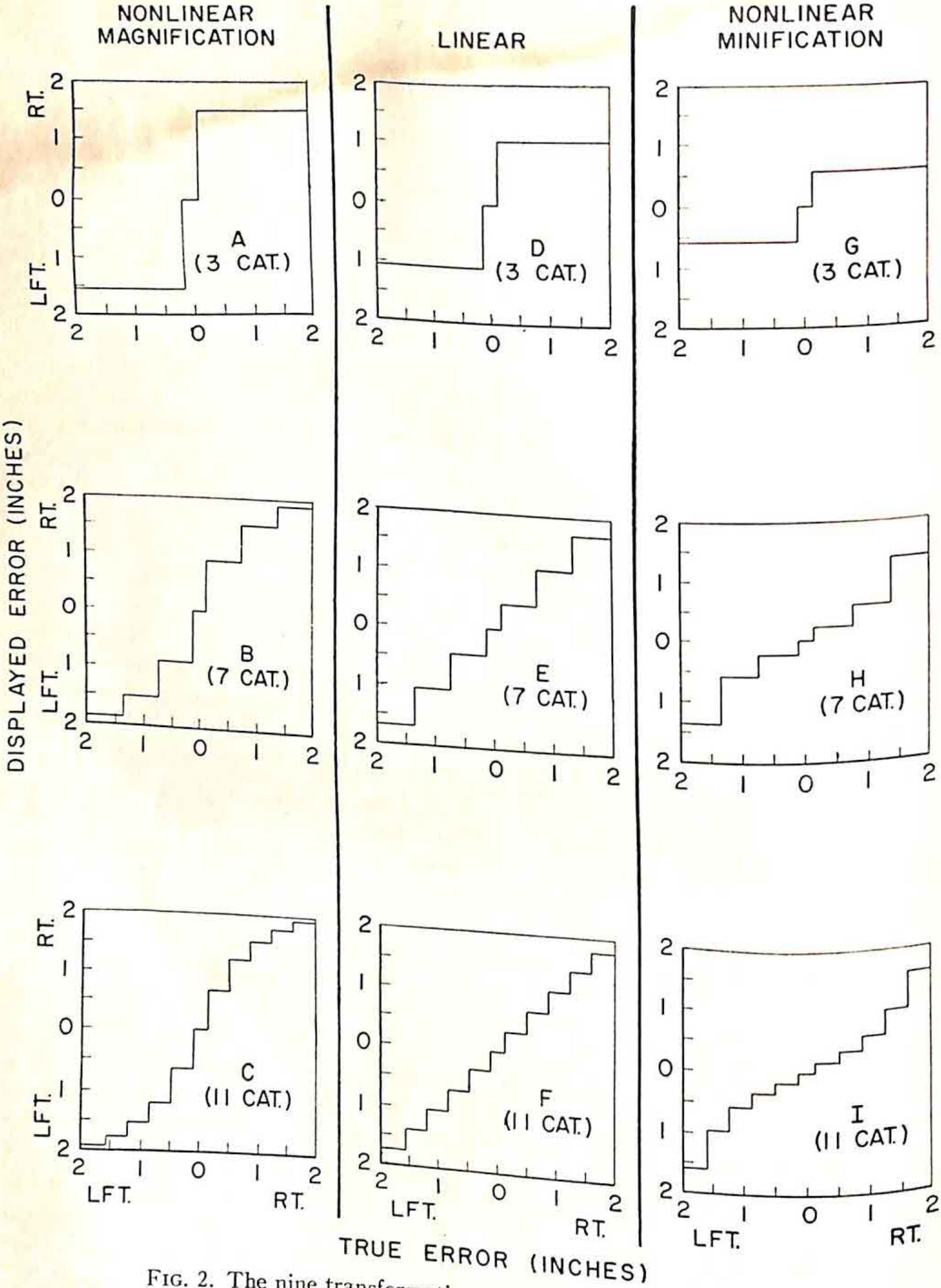


FIG. 2. The nine transformations of information investigated.

in. was displayed as zero error to the *S*. Three "nonlinear" (L) transformations were combined with each of three discrete (C) transformations (or levels of quantization) to obtain the nine informational conditions. The L transformations were:

1. Magnification (Mag)—the true error was nonlinearly magnified according to a logarithmic function (Fig. 2a, 2b, 2c).
2. Linear (Lin)—the displayed error was linearly related to the true error (Fig. 2d, 2e, 2f).

3. Minification (Min)—the true error was nonlinearly minified according to a logarithmic function (Fig. 2g, 2h, 2i).

Each of these three L transformations were quantized into 3, 7, or 11 categories. This resulted in the nine transformations (Mag-3, Mag-7, Mag-11, Lin-3, Lin-7, Lin-11, Min-3, Min-7, and Min-11).

A relatively constant testing environment was maintained by enclosing the control, the display, and *S* in an 8 × 8 × 6½ ft. sound-shielded booth. The booth was illuminated by a white 3-w. shielded bulb located behind and 3 ft. above *S*. The inside of the booth was lined with a black flocking material to eliminate potentially distracting reflections. A fan in the ceiling of the booth kept the air in circulation and effectively masked extraneous noise.

Procedure.—The 63 *S*s were divided into nine groups of 7 *S*s each. The nine groups corresponded to the nine experimental conditions of information transformations discussed in the Apparatus section. Each *S* was tested under only one of the nine informational conditions and was given 20 1-min. trials on each of 2 successive days. Rest periods of 30 sec. duration were interspersed between all trials except 10 and 11, between which a 4-min. rest was given; and, of course, about 23 hr. transpired between the end of Trial 20 and the beginning of Trial 1 on the second day. During the 30-sec. rests, *S* remained seated in the booth. During the 4-min. rest he left the booth and his activities were unrestricted.

Only seven *S*s could be tested during one day. The conditions that were run on any day were selected at random with the following restrictions: (*a*) all "linearity" conditions must be represented on each day, (*b*) none of the nine conditions could occur more than once on each day, and (*c*) over the complete experiment, each condition must occur once at each time of day (9:00, 10:00, and 11:00 A.M., and 1:00, 2:00, 3:00, and 4:00 P.M.). Unanticipated apparatus difficulties, procedural errors, and the failure of some *S*s to appear for testing at their appointed times resulted in some departure from the third restriction. In no case, however, was a deviation of more than 1 hr. permitted.

Upon the appearance of each *S*, he was seated in the booth and read a set of standardized instructions which indicated that his task was to attempt to keep the dot centered on the zero error line of the display, informed him of the procedural details (ready signals, duration of rest periods, etc.), and suggested that he would be able to track better if he made rapid and small movement of the control lever.

The *E* then left the booth and sounded a ready signal, which was a loud, but not disturbing buzzer. Three seconds later, the trial was begun. At the end of the trial, the buzzer was sounded again; 27 sec. later, the buzzer was again sounded to indicate that a trial was about to begin, the trial was started 3 sec. later, etc. No knowledge of results was given to *S* during the experiment.

RESULTS

Integrated tracking error.—The mean of the e scores for the seven *S*s who performed under each of the nine informational conditions was computed for each trial. These means are plotted for the Min, Lin, and Mag conditions in Fig. 3. Inspection and comparison of these graphs suggest that (*a*) performance improved with practice for all groups, (*b*) more improvement occurred on the first day than on the second day, (*c*) lower error is associated with a greater number of informational categories, (*d*) performance depends upon the linearity, and (*e*) especially for practiced *S*s, the reduction in e associated with an increased number of categories depends upon the nature of the linearity. These latter two relationships can be seen more clearly in Fig. 4 in which the mean error for all trials combined on the (*a*) first day, and (*b*) second day, is shown for each L condition as a function of the number of categories. The interaction between nonlinearity and number of categories is reflected in Fig. 4b by the convergence of the three nonlinearity functions as the number of categories increases.

An analysis of variance (Lindquist, 1953) was computed in order to test these apparent relationships. The results of this analysis of variance are summarized in Table 1. With the exception of the anticipated interaction between L and C, the relation-

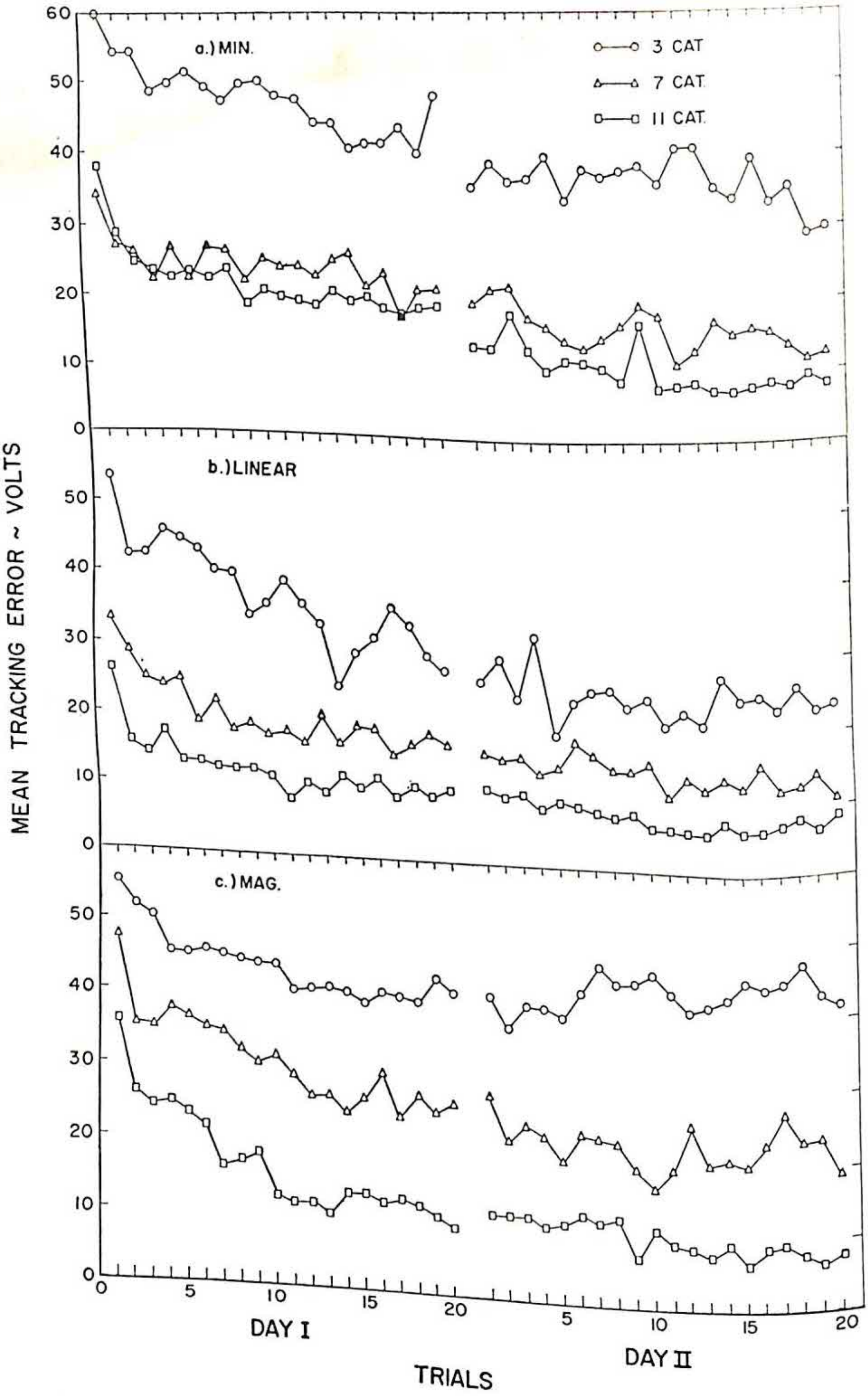


FIG. 3. Tracking error as a function of trials for the (*a*) nonlinear minification, (*b*) linear, and (*c*) nonlinear magnification conditions. (The means are plotted separately for 3, 7, and 11 categories of information. Each point is the mean of seven *S*s.)

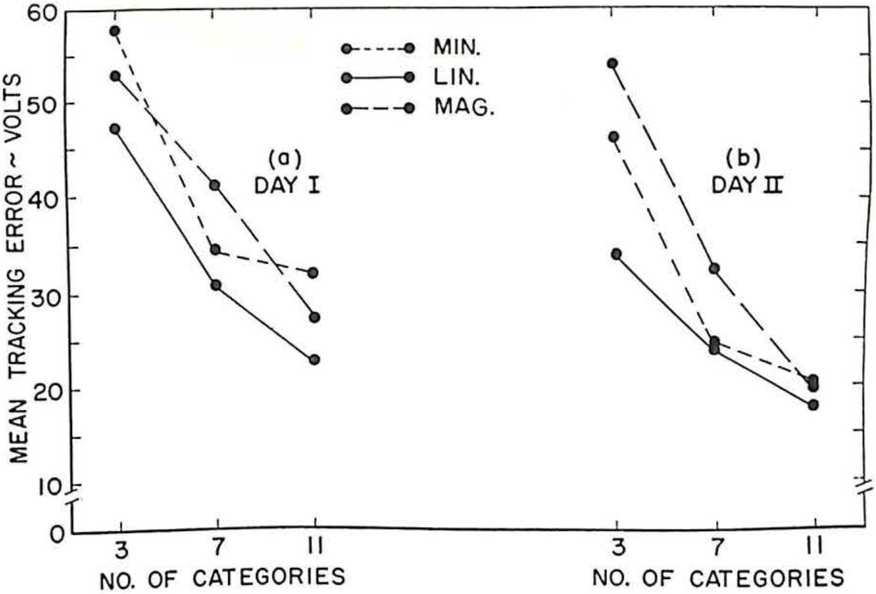


FIG. 4. Tracking error as a function of the number of informational categories for the (*a*) first, and (*b*) second day. (The means are plotted separately for the nonlinear minification, the linear, and the nonlinear magnification conditions. Each point is the mean of the 20 trials for the seven *S*s who performed under the indicated condition.)

ships expected on the basis of inspection of the data are supported; all of the following sources were significant: trials, days, nonlinearity, number of categories, and the Trials × Days interaction. To provide a closer statistical inspection of the effects of L and C after a relatively stable level of performance had been attained, the e scores for the last 10 trials were subjected to an analysis of variance; nonlinearity, number of categories, trials, and *S*s were included in the analysis. Only nonlinearity, $F(2, 54) = 4.79$,

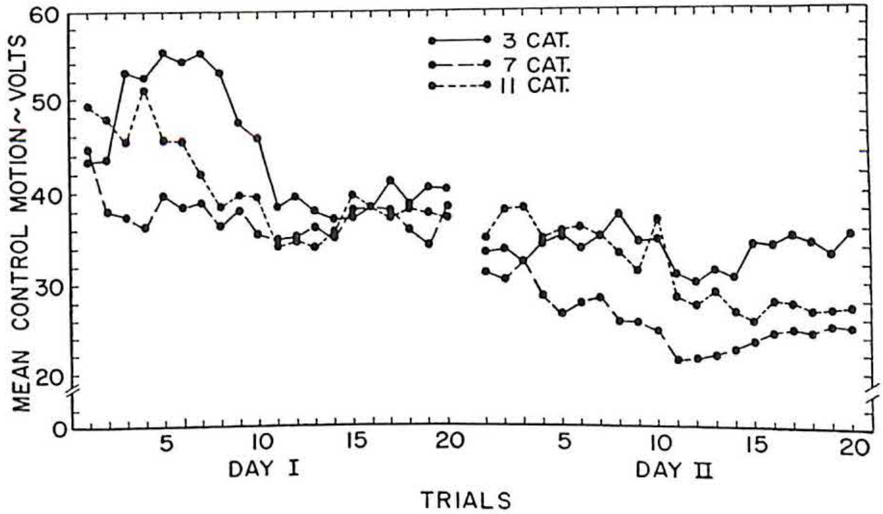


FIG. 5. Control motion as a function of trials. (The means are plotted separately for 3, 7, and 11 categories of information. Each point is the mean of 21 *S*s.)

TABLE 1

Analysis of Variance of the Error and Motion Scores

Source	*df*	Error Scores		Motion Scores	
		MS	*F*	*MS*	*F*
Between *S*s					
Nonlinearity (L)	2	3311311	4.24*	703346	2.20
Number of categories (C)	2	29215342	37.43***	298500	—
C × L	4	622066	—	416717	1.30
Error (b)	54	780595	—	319383	—
Within *S*s					
Days (D)	1	8810630	56.88***	1699934	23.48***
Trials (T)	19	221810	21.41***	33160	9.79***
L × T	38	7255	—	2550	—
L × D	2	251630	1.62	74445	1.03
C × T	38	5972	—	5029	1.49*
C × D	2	5879	—	9044	—
T × D	19	87599	11.54***	3082	—
L × C × T	76	9399	—	2675	—
L × C × D	4	263216	1.70	24702	—
L × T × D	38	10857	1.43	1985	—
C × T × D	38	10695	1.41	5792	1.75**
L × C × T × D	76	8338	1.10	2303	—
Error (w)	2106				
Error_1 (w)	54	154894		72390	
Error_2 (w)	1076	10362		3386	
Error_3 (w)	1026	7592		3314	
Grand total	2519				

* $p < .05$.
** $p < .01$.
*** $p < .001$.

$p < .05$, and number of categories, $F(2, 54) = 30.94$, $p < .001$, were significant. Neither trials nor any of the interactions approached statistical significance.

Integrated control motion.—The means of the control-motion scores for the 3-, 7-, and 11-category conditions were computed for each trial; all L conditions were combined. These means are plotted in Fig. 5. The results of an analysis of variance of the CM scores are shown in Table 1. The effects of both trial and days were significant. Unanticipated was a significant effect of the interaction of number of categories and trials, and a significant interaction among trials, days, and number of categories. The L conditions had no significant influence (either alone or in combination with any of the other variables) on the CM scores.

DISCUSSION

With regard to tracking accuracy as measured by the integrated absolute value of the untransformed tracking error, poorer performance was associated with departures from a linear display and with a smaller number of informational categories.

No scale markings other than a single index mark at the zero position were presented and the transform condition under which each *S* operated was not described or explained to him. Thus, under the Min condition the displayed information suggested that the tracking error was less than it actually was; while under the Mag condition, the reverse was true, i.e., the information provided to him implied that the tracking error was greater than it actually was. Thus, two factors were involved: (*a*) a nonlinear transformation, and (*b*) an implication that the informational display was actually linear or at least *S* was given no explicit information to indicate otherwise. The present study provides no data from which to determine the separate effects of these two factors on performance. The hypothesis that *S*s initially assume that the displayed error is equal to the true error has been proposed by Bilodeau (1953) and others (Noble & Broussard, 1955) to account for the performance of the relatively simple psychomotor task of making micrometer settings while blindfolded. The effects of the implied linearity when a nonlinear informational transformation is used could have been reduced, but probably not eliminated, by placing nonlinearly spaced scale markings on the display corresponding to the nonlinear transformation being used; but this, of course, is a separate problem and a topic for further research.

In the case of the compensatory tracking task used in this study, the nonlinear transformation of error information also could be viewed as an alteration of the controller dynamics. That is, ignoring the influence of the forcing function, a given displacement of the control does not result in a simple acceleration, but rather, in a simple acceleration *plus* the informational transformation under which *S* is performing. Furthermore, the simple acceleration is altered differently depending upon the magnitude and direction of the untransformed error.

Statistical support for the anticipated interaction between nonlinearity and number of categories (for the e scores) failed to materialize. None of the interactions involving both nonlinearity and number of categories was significant.

Figure 3, together with the statistical analysis, indicates that performance as measured by the e scores had stabilized reasonably well on the second day for all conditions. However, this does not appear to be the case when the CM scores are considered. Both the inspection of Fig. 5, as well as the lack of significance of the T × D interaction for the CM scores, suggest that the CM scores are still generally decreasing during the second day of practice.

Figure 5 also suggests that the significant C × T × D and C × T interactions for the CM scores are due largely to the 3-category condition. On Day 1 the overall slope, indicating rate of improvement, is greater for the 3-category condition than for either the 7- or the 11-category conditions. On Day 2, the 3-category condition exhibits a relatively stable level of performance, while the 7- and 11-category conditions continue to show improvement over at least the first half of the second day. It is also of interest to observe that the mean control motion on each of the final 17 trials is lower for the 7-category condition than it is for the 11-category condition.

As measured by the CM scores, the terminal level of performance under the 3-category condition on Day 2 almost had been attained by Trial 11 of the first day (Fig. 5). Little improvement is indicated through the remaining 29 trials. Exceptions to this failure to improve are observed on the trials following longer rest periods, i.e., Trial 1 on Day 2 and Trial 11 on Day 2. However, this improved level of performance is not maintained. Instead, each of these momentary improvements is followed by a gradual degradation. The CM scores under the 7- and 11-category conditions also exhibit a sizable decrease on the trials following the longer rest periods, but a degradation does not take place in all cases.

Inspection of Fig. 3 reveals that only the nonlinear minification condition (Fig. 3a) consistently shows any evidence of improving tracking accuracy, i.e., reduced e scores, on trials following the longer rest intervals. Neither the nonlinear magnification nor the linear conditions consistently exhibits such an improvement.

The improved level of performance on the postrest trials, followed by the gradual degradation, could be accounted for by appealing to the concept of inhibition; inhibition accumulates during practice and has an opportunity to dissipate during the longer rest periods; when practice is resumed, inhibition again accumulates. Regardless of the interpretation, the possibility that performance under the various conditions was differentially influenced by the particular rest intervals used in this experiment needs to be considered in applying the findings.

REFERENCES

BILODEAU, E. A. Speed of acquiring a simple motor response as a function of the systematic transformation of knowledge of results. *Amer. J. Psychol.*, 1953, 66, 409–420.

HUNT, D. P. The effect of the precision of informational feedback on human tracking performance. *Hum. Factors*, 1961, **3**, 77–85.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

NOBLE, C. E., & BROUSSARD, I. G. Effects of complex transformations of feedback upon simple instrumental behavior. *J. exp. Psychol.*, 1955, **50**, 381–386.

SCHOEFFLER, M. S. The amount of information presented as a parameter of tracking performance. *USA Med. Res. Lab. Rep.*, 1955, No. 197.

(Received June 24, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 495–496

SUPPLEMENTARY REPORTS

ACQUISITION AND REVERSAL OF A SPATIAL RESPONSE AS A FUNCTION OF SUCROSE CONCENTRATION

JAMES R. ISON

University of Rochester

Acquisition and reversal of a spatial discrimination were given to 4 groups of rats ($n = 8$) resulting from the 2 × 2 combination of 2 reinforcement magnitudes in acquisition and reversal. The reinforcement was .6 cc of either a 4% or a 20% sucrose solution. Acquisition concentration (AC) had a significant effect ($p < .05$) on acquisition performance, but no effect on reversal trials to the 1st reinforcement or trials to the reversal criterion. The latter measure was related to reversal concentration (RC, $p < .005$) and the AC × RC interaction ($p < .025$), the interaction resulting because changes in concentration from acquisition to reversal led to faster learning. The interaction suggests that part of the stimulus complex to which the spatial response is associated results from the stimulus feedback of the anticipatory goal response.

It has been concluded (Pubols, 1961) that reinforcement magnitude affects only concurrent performance and not subsequent reversal performance in spatial-discrimination learning. However, Wike and Farrow (1962) demonstrated an indirect effect of acquisition magnitude on reversal performance, which was, namely, that *S*s for which the magnitude changed from acquisition to reversal learned more rapidly than *S*s for which the magnitude remained constant. In that experiment *S*s were given a fixed number of choice and forced trials to different amounts of water reinforcement. In the present experiment *S*s were given free trials to criterion performance and the reinforcements were equal volumes of different sucrose concentrations. This manipulation of magnitude results in approximately the same consummatory time for each magnitude, in contrast to other methods in which larger magnitudes require longer consummatory times.

Method.—The *S*s were 32 naive female hooded rats, 125 days old, maintained on 23 hr. food and water deprivation. The apparatus was a glass-enclosed E maze, 8 in. high and 4 in. wide, having a gray 18-in. stem, two black 18-in. arms, and two right-angled black 8-in. goal boxes. Retrace doors were placed 6 in. inside the stem and at the entrance to each arm and each goal box. On Days 1–7 *S*s were handled and maintained on the schedule, on Day 8 they were given a single preference trial with an empty glass caster in each goal box, and on Day 9 they were given two trials in which the caster on the nonpreferred side contained the appropriate sucrose solution. Thereafter 10 trials per day were given except for Reversal Trials 1–10, which were preceded by a single trial on the acquisition schedule. All trials were noncorrected with a minimum intertrial interval of 10 min. The *S*s were randomly assigned (with the constraint that $n = 8$) to one of the cells of a 2 × 2 design, differing in acquisition-sucrose concentration (4% vs. 20%) and reversal concentration (4% vs. 20%). Each reinforcement was .6 cc of liquid. Acquisition training was continued to a criterion of 17 correct on 2 successive days and reversal training continued to a criterion of 10 consecutive correct responses.

Results.—Mean trials to criterion for each of the four groups were: Group 20:20, 21.4; Group 20:4, 22.2; Group 4:20, 26.5; Group 4:4, 24.5. Analysis of variance of these data yielded an acquisition concentration (AC) effect, $F\,(1, 28) = 4.33$, $p < .05$, but no significant differences attributable to any subsequent reversal treatment, each $F\,(1, 28) < 1.0$. It should be noted that acquisition concentration was necessarily confounded with a small (4 trials) but reliable difference in numbers of acquisition trials. However it is unlikely that such a small difference on the acquisition-trials dimension would produce the subsequent differences in reversal.

1. Mean trials to the first correct response were: Group 20:20, 4.1; Group 20:4, 6.4; Group 4:20, 3.4; Group 4:4, 3.8. Analysis of these data indicated that no treatment produced any significant effect.

2. Mean trials to criterion for each group were: Group 20:20, 11.6; Group 20:4, 15.6; Group 4:20, 9.0; Group 4:4, 29.1. Analysis of these data yielded a significant reversal concentration (RC) effect, $F\,(1, 28) = 13.17$, $p < .005$; a significant AC × RC interaction, $F\,(1, 28) = 5.88$, $p < .025$; and a nonsignificant AC effect, $F\,(1, 28) = 2.68$, $p < .20$.

3. It is also interesting to consider reversal performance after the first reversal reinforcement because only subsequent to this could the reversal treatment affect behavior. The mean trials to criterion after the first correct response were: Group 20:20, 7.5; Group 20:4, 9.2; Group 4:20, 5.6; Group 4:4, 25.3. Analysis of these data yielded these significant effects: RC, $F\,(1, 28) = 11.04$, $p < .001$; AC, $F\,(1, 28) = 4.71$, $p < .05$; and AC × RC, $F\,(1, 28) = 7.76$, $p < .01$.

Discussion.—These results agree with those previously obtained by Wike and Farrow (1962) in demonstrating that although the major determiner of reversal performance is reversal reinforcement, there exists a reliable effect of changes in magnitude from acquisition to reversal. The occurrence of this effect under two quite different procedures and reinforcement conditions indicates that it has reasonable generality. The interaction suggests that a significant portion of the stimulus complex to which the spatial response is associated arises as feedback from the anticipatory goal response, so that with changes in the goal response there is a greater generalization decrement of the now erroneous response. Because of the major procedural differences between the Wike and Farrow experiment and the present one, it is likely that the previous failure to observe the effect (Pubols, 1961) was due not to procedural differences but to reinforcement magnitudes being not sufficiently disparate.

REFERENCES

PUBOLS, B. H. The acquisition and reversal of a position habit as a function of incentive magnitude. *J. comp. physiol. Psychol.*, 1961, **54**, 94–97.

WIKE, E. L., & FARROW, B. J. The effects of magnitude of water reward on selective learning and habit reversal. *J. comp. physiol. Psychol.*, 1962, **55**, 1024–1028.

(Received June 13, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 496–498

EFFECT OF A READY SIGNAL ON THE LATENCY OF VOLUNTARY RESPONSES IN EYELID CONDITIONING[1]

KENNETH P. GOODRICH[2]

University of Pennsylvania

60 trials with a ready signal and 60 trials without a ready signal were presented in a quasi-random sequence to each of 20 Ss. Results showed that significantly more short-latency "voluntary" responses occurred on trials with a ready signal.

After comparing the data from their own experiment with those from a study by Spence and Ross (1959), Hartman and Ross (1961) concluded that the presence of a ready signal in an eyelid conditioning experiment probably results in shorter latencies of the so-called "voluntary" (V) responses. Consistent with this conclusion is the fact that a similar effect is produced when *S*s are instructed to blink to the CS (Hartman, Grant, & Ross, 1960). Clearly, however, the evidence from conditioning experiments is not conclusive. As Hartman and Ross point out, their interpretation rested on a comparison of two groups of *S*s in two separate experiments which differed in several other respects besides the presence and absence of a ready signal.

In the present experiment each *S* was run both with and without a ready signal. This procedure controls for individual differences as well as for possible confounding of extraneous variables with the presence and absence of the ready signal as may have occurred in the earlier work.

Method.—Twenty undergraduates from the University of Pennsylvania served as *S*s, 6 women and 14 men. Three of the women were from the introductory psychology course; the rest had had no psychology courses.

The apparatus and general procedure were essentially like those of other recent eyelid studies (e.g., Spence & Ross, 1959). The CS was an increase in illumination of a 2.5-in. diameter translucent disk from .03 to 1.73 mL. The CS-US interval was 500 msec.; the UCS, a 5-psi air puff, was turned on for 50 msec. and terminated with the CS.

---

[1] This research was supported by NSF Grant GB-203.

[2] Now at the Regional Primate Research Center, University of Wisconsin.

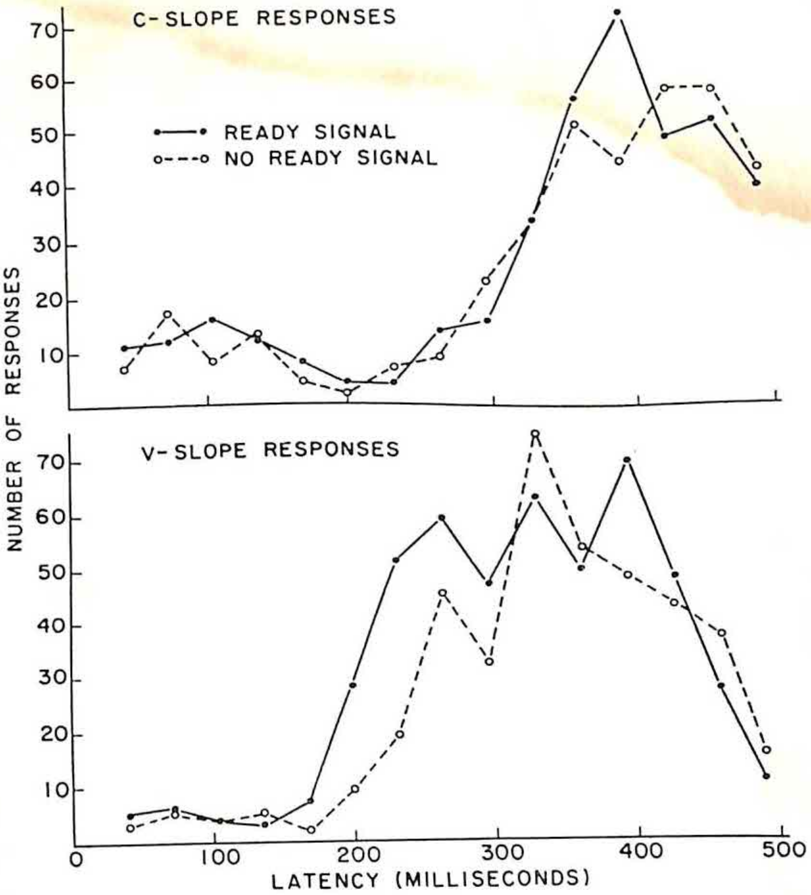


FIG. 1. Latency distributions of "nonvoluntary" (top) and "voluntary" (bottom) responses on trials with and without a ready signal.

All *S*s received 120 conditioning trials. A ready signal occurred on 60 of these. It consisted of a 400-msec. burst of white noise which preceded the onset of the CS by a quasi-random 2-, 3-, or 4-sec. interval. On the other 60 trials there was no ready signal. The two kinds of trials came in a scrambled order, five of each kind in each successive block of 10 trials.

The instructions were essentially the same as those used by Spence and Ross (1959), except that *S* was asked to look at the disk at all times (as in an experiment without a ready signal) and to blink immediately when he heard the ready signal (as in an experiment with a ready signal).

Results and discussion.—Classification of responses as "voluntary" (V) or "nonvoluntary" (C) was carried out according to the procedure suggested by Hartman and Ross (1961). The maximum slope of each response was expressed as a proportion of the maximum slope of *S*'s UCR, and if this value was as great as .40 the response was classified as a V. A response with a slope less steep than .40 was classified as a C.

In Fig. 1 are the latency distributions for V and C responses both with and without a ready signal. The ordinate in each case is the total number of responses in the appropriate latency interval across both *S*s and trials. Only the first response on any trial was included. The data in this figure clearly corroborate the data discussed by Hartman and Ross (1961). The presence of the ready signal resulted in V responses of shorter latency than did the absence of the signal. This effect was confined to the region between 150 and 300 msec.

The results of several analyses attest to the reliability of the difference in number of short-latency V responses between the two ready-signal conditions. Using the median latencies of V responses under the two conditions for each *S*, a Wilcoxon test showed that latencies were shorter on trials preceded by a ready signal ($p < .025$, one-tailed test). The same conclusion is indicated by an analysis of the number of V responses for each *S* which occurred in the interval 150–300 msec. following the CS. Using as a measure either this number or this number as a fraction of all the V responses after 150 msec. under the particular ready signal condition, a sign test showed that significantly more V responses occurred in the 150–300 msec. interval when the ready signal was used ($p = .004$, one-tailed test).

Several other features of the data are worth noting here. First, the distribution of C responses was unaffected by presence or absence of the ready signal. Second, more responses were made on trials when a ready signal was present than when it was absent. Third, over 60% of the V-slope responses occurred after 300 msec. even with a ready signal, a finding which corroborates data from our laboratory obtained when a ready signal occurred on all trials and contrasts sharply with the figure of 17% reported by Spence and Ross (1959). Fourth, responses prior to 150 msec., commonly said to be "original" responses to the CS, are mostly shallow in slope and thus unlike the "original" response to the UCS.

REFERENCES

HARTMAN, T. F., GRANT, D. A., & ROSS, L. E. An investigation of the latency of "instructed voluntary" eyelid responses. *Psychol. Rep.*, 1960, **7**, 305–311.

HARTMAN, T. F., & ROSS, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *J. exp. Psychol.*, 1961, **61**, 334–338.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

(Received June 27, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 5, 498–500

LONG-TERM REMINISCENCE IN THE PURSUIT-ROTOR HABIT

JEFFERSON M. KOONCE, DAVIS J. CHAMBLISS, AND ARTHUR L. IRION
Tulane University

Reminiscence was measured on a pursuit-rotor task following 8 different rest intervals; 0 min., 10 min., and 1, 7, 35, 70, 175, 365, and 730 days. In all, 130 Ss were tested. Using the no-rest group as a control, reminiscence was obtained in all of the other conditions. There was a regular, but nonsignificant tendency for amount of reminiscence and amount of warming-up effect to decrease with the longer intervals of rest.

Although a considerable amount of investigative effort has been devoted to determining the length of rest that may produce maximum reminiscence gains in motor-learning situations (Ammons, 1947; Irion, 1949; Kimble & Horenstein, 1948), relatively little research has been directed toward determining how long the reminiscence effect may persist. Melton's (1941) study indicated that reminiscence gains could be detected following a rest interval of 6 wk., and Jahnke and Duncan's (1956) systematic investigation revealed little decrease in the reminiscence effect over rest intervals of up to 4 wk. On the other hand, Bell (1950) failed to obtain reminiscence after a period of 1 yr., although, in this study, prerest training was given under conditions of relatively spaced practice—a condition unfavorable to the production of large amounts of reminiscence under any condition of rest. The previous findings suggest that reminiscence gains and warming-up effects increase as rests are lengthened from 10 min. to 24 hr. and that, with longer rests, both effects diminish.

The present study was designed to yield large amounts of reminiscence with optimum lengths of rest and to study the persistence of the reminiscence gains over an extended period of time.

Method.—The *S*s were 130 right-handed, male, undergraduate students in Tulane University who were registered in introductory psychology classes during the fall semester of 1960–61. Handedness was determined by a questionnaire that asked *S* with which hand he performed a variety of common tasks. Any "left-handed" answer served to disqualify *S*. Because it was desired to use rest intervals of more than a year, senior students were excluded. All *S*s volunteered to participate in the experiment. In all, 200 *S*s were given initial practice on the learning task although not all of these *S*s were able to complete the postrest phase of the experiment.

Apparatus.—A unit of three standard, 60 rpm, United States Air Force pursuit rotors, mounted, together with a control panel, on a USAF psychomotor table was used. Time on target was measured by means of six Standard Electric timers that were calibrated in .001 min. All scores reported in this experiment are given in thousandths of a minute.

Prerest practice.—The *S*s reported to the laboratory in groups of three and were assigned to individual rotors at random. Following instruction in the task, they were given 5 min. of continuous practice. During this practice, scores were taken at 30-sec. intervals, and these 30-sec. intervals or measurement periods were arbitrarily designated as "trials." At the end of practice, all *S*s were dismissed (with the exception of *S*s who had been assigned to the no-rest and 10-min. rest conditions and who were given their postrest trials either immediately or after 10 min.). No *S* was informed that he would be asked to return for a second experimental session.

Rest intervals.—Length of rest constituted the independent variable. Eight different rest intervals (10 min., and 1, 7, 35, 70, 175, 365, and 730 days) were used. In addition, a ninth group was given 10 min. of continuous practice and served as the no-rest, control condition.

Some problems of sampling should be mentioned briefly. Since *S*s were given initial training in groups of three, assignment of *S*s to experimental conditions was not completely random. Groups of three *S*s each were randomly assigned to the no-rest and 10-min. rest conditions. The remaining *S*s were put in random order and were called back for the postrest training in that order. Because relatively long retention intervals were being used, it was felt that there might be a systematic difference between *S*s in the no-rest and 10-min. rest groups and those in the other conditions in terms of willingness to return for a second experimental session, since *S*s in the no-rest and 10-min. rest groups, if they appeared for prerest practice, were "captive" *S*s for the postrest trials whereas this was not true for *S*s in the other groups. In order to control for this source of bias, *S*s in the no-rest and 10-min. rest groups were asked to return to the laboratory for a second, "dummy," test approximately 5 wk. following their first session. Those *S*s who failed to keep the second appointment were dropped from the experiment. Another possible source of bias should also be noted. The *S*s who were given rest periods of 70 days or less were given their prerest and their postrest trials within a single semester, but for *S*s who received rest periods longer than this there was an opportunity for *S*s to drop (or to be dropped) from school. This form of selection may have had some effect upon the results, although a comparison of the prerest scores between the appropriate groups does not reveal a significant difference.

Finally, it should be noted that there were 15 *S*s in each group except the 730-day group which contained 10 *S*s. The lower *N* in the 730-day group reflects the fact that not enough *S*s were run in 1960 to compensate for the attrition suffered by *S*s over a 2-yr. period.

Postrest practice.—Relearning consisted of 5 min. of additional continuous practice divided, as before, into 10 30-sec. "trials" or measurement periods. Postrest practice for each *S* was given on the same rotor as was his prerest practice. An effort was made to run *S*s in groups of three on their postrest trials, but scheduling difficulties prevented this in most instances. Some *S*s received their postrest trials alone, but most of them worked with one or two other *S*s who could be scheduled at the same time.

Results.—The trial means for the prerest performances in all groups are presented in Table 1 and the trial means for the postrest performances are given in Table 2.

It will be recalled that a number of earlier studies have shown that postrest performance increases rapidly as a function of length of rest as rest-interval length increases up to

TABLE 1

TRIAL MEANS IN PREREST PRACTICE

Cond.	Trials										
	1	2	3	4	5	6	7	8	9	10	*M*
10 min.	62	67	75	78	97	110	109	130	117	124	97
1 day	74	79	94	103	102	121	129	119	136	124	108
7 days	68	86	90	99	105	113	109	122	119	145	106
35 days	80	79	83	87	86	93	103	113	133	127	98
70 days	62	70	81	74	89	97	103	113	117	109	92
175 days	60	69	71	91	98	106	110	124	107	113	95
365 days	74	74	100	109	111	129	121	132	146	143	114
730 days	60	74	85	91	89	106	109	121	117	124	98
Mean of experimental conditions	68	75	85	92	98	110	112	122	124	126	101
No-Rest (Control)	66	62	80	100	95	110	109	115	120	140	100

Note.—All scores are in .001 min. and are rounded to the nearest .001 min. Means may be converted to percentage time-on-target scores by multiplying by .2.

TABLE 2

TRIAL MEANS IN POSTREST PRACTICE

Cond.	Trials										
	11	12	13	14	15	16	17	18	19	20	*M*
10 min.	224	269	275	259	260	247	249	232	246	234	250
1 day	212	256	266	254	248	270	264	262	246	252	253
7 days	212	250	273	260	276	248	254	252	248	252	252
35 days	211	247	250	245	238	241	242	240	234	231	238
70 days	174	202	222	219	231	223	229	222	209	220	215
175 days	169	188	187	186	205	199	202	194	193	202	192
365 days	180	214	213	232	227	227	224	237	220	208	218
730 days	160	189	196	209	197	213	210	214	216	221	203
Mean of experimental conditions	194	228	237	234	237	234	235	232	227	228	229
No-Rest (Control)	130	136	130	135	127	143	136	156	158	162	141

Note.—All scores are in .001 min. and are rounded to the nearest .001 min. Means may be converted to percentage time-on-target scores by multiplying by .2.

about 10 min. Since 10 min. was the shortest interval of rest used in the present experiment, it might be expected that postrest performance would decrease as a function of still longer rests. Although reminiscence gains appear to be remarkably impervious to decay, the data in Table 2 indicate that Trial 11 scores and the total postrest scores do tend to decrease as length of rest is increased up to 730 days. If the gain in score between the first and second postrest trial is taken as an indication of amount of the warming-up effect, it will be noted that this warming-up effect also tends to decrease as a function of length of rest (it having been already established that, with shorter rests, the warming-up effect increases as a function of length of rest). However, these trends need to be qualified in terms of the results of the statistical analysis reported below.

The pre- and postrest scores were analyzed by conventional analysis-of-variance techniques. The results of these analyses may be summarized under the following five headings:

1. There were no significant differences between all groups in prerest performance either with respect to conditions or Conditions × Trials interaction.

2. A significant difference at the 1% level was found between conditions in postrest performance, $F(8, 121) = 4.43$. This difference is attributable to the differences between the no-rest (control) condition and the other conditions, however, because:

3. When the no-rest condition is excluded from the analysis of postrest performance, there is no significant difference between conditions.

4. The Conditions × Trials interaction for postrest performance is significant at the 5% level, $F(72, 1089) = 1.413$, but, again:

5. When the no-rest, control condition is excluded from the analysis of postrest performance, the Conditions × Trials interaction term is not significant.

Discussion.—The experimental results appear to be more or less self-explanatory. As is usually found in the pursuit-rotor situation, large amounts of reminiscence were obtained. The fact that the postrest scores for the 10-min., 1-day, and 7-day groups cluster together bears out the usual finding that reminiscence quickly rises to a stable, maximum value as a function of length of rest. The high degree of permanence of the reminiscence gains is, perhaps, somewhat surprising. Even though the trend of the results would indicate that, if sufficiently long rests were to be employed (and, assuming adequate longevity of *S*s), the reminiscence effect would probably decrease to zero, the gains appear to be remarkably impervious to decay. The trend of the results also gives general support to the conclusion of Jahnke and Duncan (1956) that warming-up effects at first increase and then decrease as a function of length of rest. The relative regularity of the data supports such conclusions even though the differences between means and the Groups × Trials interaction among the experimental groups are not statistically significant.

REFERENCES

AMMONS, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.

BELL, H. M. Retention of pursuit rotor skill after one year. *J. exp. Psychol.*, 1950, **40**, 648–649.

IRION, A. L. Reminiscence in pursuit-rotor learning as a function of length of rest and of amount of pre-rest practice. *J. exp. Psychol.*, 1949, **39**, 492–499.

JAHNKE, J. C., & DUNCAN, C. P. Reminiscence and forgetting in motor learning after extended rest intervals. *J. exp. Psychol.*, 1956, **52**, 273–282.

KIMBLE, G. A., & HORENSTEIN, B. R. Reminiscence in motor learning as a function of length of interpolated rest. *J. exp. Psychol.*, 1948, **38**, 239–244.

MELTON, A. W. The effect of rest pauses on the acquisition of the pursuitmeter habit. *Psychol. Bull.*, 1941, **38**, 719. (Abstract)

(Received July 10, 1963)

Journal of
Experimental Psychology

Vol. 67, No. 6 June 1964

SYSTEMATIC STUDY OF END ANCHORING AND CENTRAL TENDENCY OF JUDGMENT

DONALD M. JOHNSON and CALVIN R. KING, Jr.

Michigan State University

The stimuli were slides of 36 green and blue dots, varying from 0 green (36 blue) to 36 green (0 blue) dots. From these 37 slides 7 sets of 11 slides each were drawn and presented for judgment on a scale of 0 to 10. Anchoring occurred where the end slide was salient, e.g., monochromatic, in accordance with the subjective-standard hypothesis. Anchoring of the other slides could be interpreted as due to the effect of the monochromatic slide, even when it was not presented. No anchoring occurred when neither end was salient. Regression toward the mean, i.e., central tendency, occurred at the unanchored end, or ends.

When an *S* repeatedly judges each of an ordered series of stimuli on an ordered scale of judgment, his activity can reasonably be described as one of matching the scale of judgment he is instructed to use to the series of stimuli he is given. Usually the matching is successful in that the scale of judgment corresponds roughly to the series of stimuli. The scale is anchored, in a general way, to all the stimuli, but many experiments have shown that certain response categories are anchored to certain stimuli in a special way. In judgments of inclinations of lines, for example, the horizontal and vertical axes are salient features to which judgment scales are easily anchored (Onley & Volkmann, 1958). The most frequently reported anchoring effect is end anchoring, in which the ends of the response scale are linked to the ends of the stimulus series, with the result that judgments of the end stimuli are made with fewer errors and less variability than judgments of the middle stimuli (Hunt, 1941; Needham, 1935; Volkmann, 1951). The effect spreads inward, in the case of an ordered series, hence it is usually assumed that the stimuli near the ends are judged in relation to the end stimuli (Eriksen & Hake, 1957). Eriksen and Hake eliminated the ends of the stimulus series by the use of colored papers in an endless continuum and found that their *S*s selected other salient features, such as the ends of the response scale, as anchors. Hence they favor the "subjective-standard hypothesis," i.e., that *S* selects any salient feature that will help him stabilize his judgments, and interpret end anchoring as a special case illustrating this hypothesis.

If the subjective-standard hypothesis is accepted, the next step is to make it more specific. The present experiment attempts to specify the conditions under which an end stimulus will be selected as a standard to which the response scale is anchored. Certainly the end of a stimulus series is not always identified as such. It is a reasonable hypothesis, in the light of the experiments mentioned above, that an end stimulus acts as an anchor when and only when it has some salient property that identifies it as an end beyond which there are no other stimuli to be judged.

Investigation of this hypothesis requires stimulus series in which the salience of the end stimuli can be varied, and the material used by Philip (1947) is convenient for this purpose. He used 11 stimuli of 36 green and blue dots, ranging from 13 green dots (with 23 blue) to 23 green dots (13 blue). Complete extension of the range from 0 green dots (36 blue) to 36 green dots (0 blue) yields 37 stimulus patterns from which various sets can be systematically drawn to fit the requirements of the present experiment.

Such stimulus material has several advantages. Stimuli of 0 green dots (all blue) and 36 green dots are monochromatic and thus easily identified as ends beyond which no other stimuli can be expected. Stimuli adjacent to the ends have one or two dots contrasting to the larger mass of the other color, hence there is a gradient of salience at each end of the complete series. The complete range is symmetrical in that it extends from no green dots, through half of each, to no blue dots and thus minimizes asymmetries connected with Weber's law. Furthermore, it is possible to use a set in the middle range, from 13 to 23 green dots, neither end of which is salient, and predict that no anchoring will occur.

Anchoring is demonstrated by a decrease in variability and errors of judgment of the anchoring stimulus. Since the error measure has certain statistical advantages, the number of response categories was made equal to the number of stimuli in each set. Each stimulus series was composed of 11 stimuli and each response scale was defined as 0 to 10.

METHOD

The complete series of stimuli was composed of 37 1-in. slides of blue and green dots in random patterns. The slides were coded 0 to 36 in accordance with the number of green dots. Seven sets of 11 slides each were drawn so as to vary the salience of the end stimuli of a set. Set 1, composed of Slides 0–10, has a salient end, Slide 0. Set 2, composed of Slides 1–11, also has a salient end, Slide 1. Set 3, of Slides 2–12, has a somewhat less salient end, Slide 2. In Set 4, of Slides 13–23, neither end is salient. Set 5, composed of Slides 24–34, Set 6, of Slides 25–35, and Set 7, of Slides 26–36, are the complements of Sets 3, 2, and 1, with salient stimuli at the other end of the series.

The *S*s were instructed to judge the slides on a scale of 0–10. One group judged on a scale of greenness, with 10 as maximum greenness, and another judged on a scale of blueness, with 10 as maximum blueness, in order to balance out any unforeseen peculiarities in the fit of the scale to the stimuli.

Seventy *S*s were recruited from elementary psychology classes and randomly assigned to 14 groups of 5 each, for the seven stimulus sets judged on the two scales. The 5 *S*s in each group were run as a group. The only criterion for recruitment was a request that no color-blind students participate.

The slides were projected by a Viewlex projector with an Alphax shutter for 1 sec., with an interslide interval of approximately 8 sec. The screen was about 20 ft. from the *S*s in a semidark room. The slides were shuffled between sets, and the orientation of the slides in the projector was varied to minimize extraneous cues. Each group judged the set of 11 slides twice for practice then six times for the record. Total time for a group was about 40 min.

RESULTS

The results were first analyzed in respect to errors. Since each set had 11 stimuli and the response scale had 11 numbers, an error was defined as any deviation from agreement. The first question was whether the data of the two scales, blueness and greenness, could be combined. Table 1, summarizing an analysis of variance of the errors, shows no difference between the scales in this respect, hence the blueness scale was reflected and the data combined with the data from the greenness scale in subsequent analyses. Table 1 also shows, as expected, a large variance between sets in number of errors.

TABLE 1

OVERALL ANALYSIS OF VARIANCE OF ERRORS

Source	df	F
Sets	6	18.99**
Scales	1	0.39
Set × Sc	6	1.76
Error (MS)	140	(36.98)

** $p < .01$.

Figure 1 gives an overall view of the variability of the judgments, with both scales combined, in the form of quartiles of the distributions of the 60 judgments of each slide. Any judgment that does not fall on the diagonal line in these plots is an error. It is obvious that the stimuli at the salient end of Set 1 and the salient end of Set 7 were judged with few errors and low variability. This anchoring effect is not limited to the anchoring stimulus but spreads to adjacent stimuli, decreasing with distance from the anchoring stimulus. The same statements can be made for Sets 2 and 6. This end anchoring also appears in Sets 3 and 5 but to a lesser extent. Set 4, in which the ends are not salient, shows no anchoring. All slides appear to be judged with equal variability.

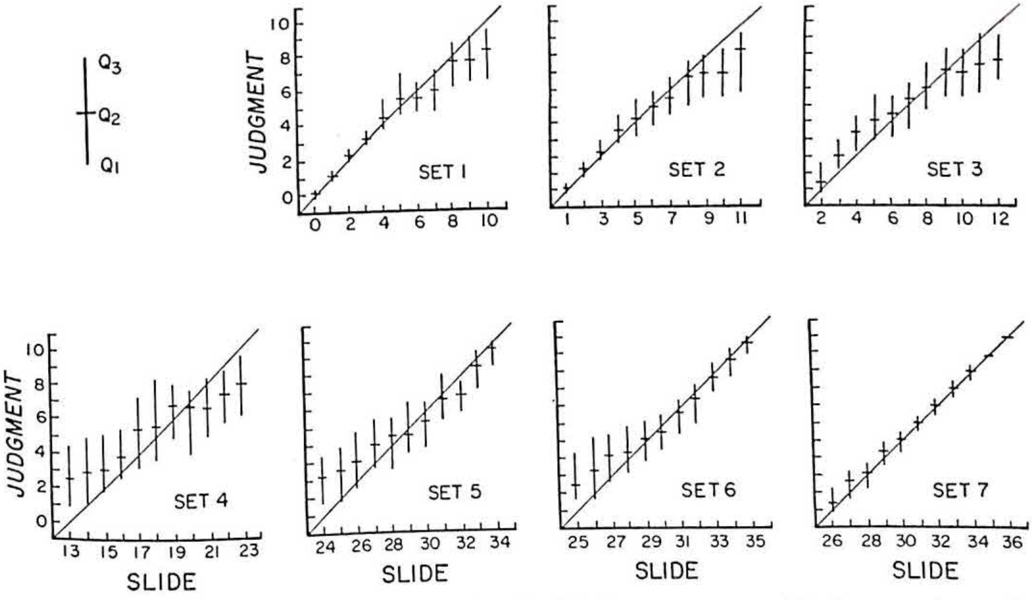


FIG. 1. Distributions of judgments of each slide in seven sets. (Medians and quartiles are shown for each distribution of 60 judgments. Correct judgments fall on the diagonal line. The blueness scale has been reflected to accord with the greenness scale.)

TABLE 2

Analysis of Variance of Errors for Each Set

Set	Source	*df*	*MS*	*F*
1	Slides	10	29.876	22.858**
	*S*s	9	9.152	
	Sl × *S*s	90	1.307	
2	Slides	10	21.216	10.791**
	*S*s	9	5.206	
	Sl × *S*s	90	1.966	
3	Slides	10	4.307	3.059**
	*S*s	9	2.782	
	Sl × *S*s	90	1.408	
4	Slides	10	0.797	1.014
	*S*s	9	0.893	
	Sl × *S*s	90	0.786	
5	Slides	10	4.222	3.707**
	*S*s	9	2.387	
	Sl × *S*s	90	1.139	
6	Slides	10	7.910	4.287**
	*S*s	9	13.998	
	Sl × *S*s	90	1.845	
7	Slides	10	22.005	20.799**
	*S*s	9	9.563	
	Sl × *S*s	90	1.058	

** $p < .01$.

As a check on the trends that appear in Fig. 1 an analysis of variance of the errors was carried out for each set. The results, shown in Table 2, are clear-cut. The variance between slides in Set 4 is not significant. The variance between slides is significant in the other sets and is very large in Sets 1 and 7, each of which has a highly salient monochromatic end.

Figure 2 has been prepared to show in more detail what happens to errors near the anchoring stimulus. Since the anchoring effects at the all-green end and at the all-blue end are similar, the error curves have been folded over so that the monochromatic stimuli in corresponding pairs coincide and the base line thus represents an increase, beginning at zero, in the number of blue or green dots. It is clear from Fig. 2 that the greater accuracy of judgment of the sets with the better anchors extends throughout the set. It also appears that Slides 1 and 35, with one contrasting dot, are judged more accurately in Sets 1 and 7 than in Sets 2 and 6 where they are the end stimuli—though this difference is not quite significant. Likewise, Slides 2 and 34, with two contrasting dots, are judged more accurately in Sets 1 and 7 than in Sets 3 and 5 where they are the end stimuli—and this difference is highly significant.

The relation between end anchoring and the central tendency of judgment is shown in Table 3 by a tabulation of the frequency of judgments in the end categories, 0 and 10, and in the middle category, 5. If all categories were used equally often, these frequencies would be 60. Some central tendency appears in all sets in that Categories 0 and 10 were used less often than Category 5. But this tendency is less

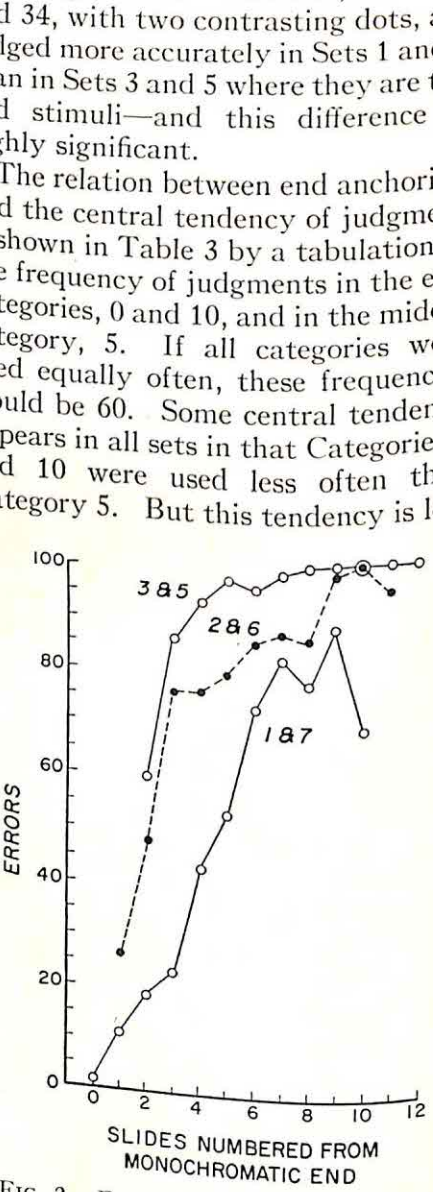


Fig. 2. Errors in judgment of each slide arranged in pairs of sets by distance from the monochromatic slide. (Each point represents the number of errors in 120 judgments.)

TABLE 3

Frequency of Middle and End Judgments and Correlations between Stimuli and Judgments for Each Set

Set	Frequency of Judgment			r
	0	5	10	
1	65	76	36	.86
2	59	75	23	.78
3	41	73	33	.71
4	39	70	45	.50
5	41	70	50	.69
6	41	68	50	.74
7	54	68	60	.96

apparent at the anchored ends of the anchored scales. A general statement that would apply to all these data, including the unanchored Set 4, is that the categories at unanchored ends of scales of judgment are used with less than average frequency.

Since the central tendency of judgment has been interpreted as a manifestation of the lack of perfect correlation between judgments and stimuli (Guilford, 1954; Johnson, 1955), the correlation coefficients are also shown in Table 3. These correlations are consistent with the *F* ratios of Table 2 in showing close agreement between judgments and stimuli for those sets in which strong anchoring effects were predicted. With the exception of Set 3, there is an inverse relation between the magnitude of the correlation coefficients and the amount of central tendency.

Discussion

The data for Set 4 demonstrate that end anchoring does not always occur. This set of stimuli was the same as the other sets in which end anchoring did occur except that neither end of Set 4 was salient. The results agree with the subjective-standard hypothesis and make it more specific.

The monochromatic stimuli, Slides 0 and 36, were the best anchors in this experiment. Anchoring effects were progressively smaller when Slides 1 and 35, with one dot of a different color, and Slides 2 and 34, with two dots of a different color, were the end stimuli. The obvious interpretation is that a slide with one contrasting dot out of 36 is less salient than a monochromatic slide and that a slide with two contrasting dots is less salient than a slide with one. The evidence does not refute this interpretation, but the comparisons graphed in Fig. 2 suggest the possibility that the monochromatic slides were the subjective standards selected as anchoring stimuli for all sets, even for the sets in which these slides did not appear. Since *S* in an absolute-judgment experiment frequently anticipates stimuli that he has not seen, an end stimulus beyond which no other can be anticipated would be exceptionally salient and would thus supply a strong anchor for one end of the scale of judgment. Blumenfeld (1931) noted the anchoring effect of an imaginary straight line on judgments of the irregularity of scattered points, and Hunt and Volkmann (1937) demonstrated the effect of "the most pleasant color you can think of" on judgments of the pleasantness of colors. Thus it is reasonable that the anchoring effects of end stimuli with one or two contrasting dots are due to their close relations to the monochromatic stimuli, even if the latter are not actually seen. Since this is a secondary effect, it would not be as strong as the primary effect of the visible monochromatic end stimuli. This interpretation would also account for the fact that Slides 2 and 34 were judged more accurately when they were in the same set with Slides 0 and 36 than when they were the end stimuli.

The central tendency of judgment has two manifestations: The middle judgment categories are used more often than the end categories, and stimuli of low value are overestimated while stimuli of high value are underestimated. Both effects reflect the lack of perfect correlation between judgments and stimuli, and any condition that makes it easier for *S* to associate judgments consistently with stimuli will raise the correlation and reduce the central tendency. One such condition is end anchoring, which ties one end of the judgment scale to one end of the stimulus series. When there is no anchor, as in Set 4, the correlation is low and the regression of judgments on stimuli deviates from the diagonal line of $r = 1.00$. A straight-line fit to the medians shown in the graph for Set 4 would approximate the regression line for $r = .50$. Thus the median judgment of the stimulus at one end was 2.25 rather than 0, and of the stimulus at the other end 7.50 rather than 10. Correspondingly, the end-judgment categories, 0 and 10, were used only 39 and 45 times, rather than 60. When one end of the scale was well anchored, this regression toward the middle did not appear at the anchored end. But the graphs for Sets 1, 2, 3, 5, and 6 demonstrate that some regression toward the center of the response scale may occur at the unanchored end, and the data of Table 3 demonstrate the same effect in another form.

It is a safe generalization that end anchoring reduces the central tendency. At the unanchored end, or ends, of a scale of judgment the judgments regress toward the mean judgment and the end categories are used with less than average frequency.

REFERENCES

BLUMENFELD, W. Urteil und Beurteilung. *Arch. ges. Psychol.*, 1931, No. 3.

ERIKSEN, C. W., & HAKE, H. W. Anchor effects in absolute judgments. *J. exp. Psychol.*, 1957, **53**, 132–138.

GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

HUNT, W. A. Anchoring effects in judgment. *Amer. J. Psychol.*, 1941, **54**, 395–403.

HUNT, W. A., & VOLKMANN, J. The anchoring of an affective scale. *Amer. J. Psychol.*, 1937, **49**, 88–92.

JOHNSON, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

NEEDHAM, J. G. Rate of presentation in the method of single stimuli. *Amer. J. Psychol.*, 1935, **47**, 275–284.

ONLEY, J. W., & VOLKMANN, J. The visual perception of perpendicularity. *Amer. J. Psychol.*, 1958, **71**, 504–516.

PHILIP, B. R. The relationship of exposure time and accuracy in a perceptual task. *J. exp. Psychol.*, 1947, **37**, 178–186.

VOLKMANN, J. Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951.

(Received July 19, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 507–513

RELATIVE SIZE IN ISOLATION AS A STIMULUS FOR RELATIVE PERCEIVED DISTANCE[1]

WILLIAM EPSTEIN AND STEPHEN S. BARATZ

University of Kansas

2 experiments examined the perceived relative distance produced by relative size in isolation. The hypothesis was that S assumes identity of physical size and interprets the difference in apparent size as a difference in distance. Experiment I examined this hypothesis by observing the effect of experimentally trained size assumptions on subsequent relative distance judgments. Experiment II used pairs of familiar objects which had known, normally invariant physical sizes. By varying the actual physical sizes of the standards it was possible to study the hypothesized effect of size assumptions on relative perceived distance. The results of Experiment I did not agree with the hypothesis. Experiment II did confirm the hypothesis. Both experiments were discussed in the context of the assumptive hypothesis.

Relative size is a stimulus correlate of perceived relative distance. The object which subtends the smaller visual angle appears more distant. This is the case when all other stimulus correlates, e.g., motion perspective, are eliminated, as well as under unrestricted conditions of viewing. One explanation of the effects of relative size under restricted conditions of observation is suggested in the following definition of the relative-size cue:

> For a given ratio of visual angles, assuming the ratio of physical sizes determines the ratio of distances. If the assumed sizes are equal, the object subtending the larger visual angle appears nearer and the one subtending the small appears farther [Ittelson, 1960, p. 69].

This definition proposes that the effect of relative size is mediated by the subjective assumption of identity. Ames (1946) has expressed this view succintly: "You presume identity of size . . . from similarity, and use the appearance of difference (in size) as an indication of distance [p. 46]." Two experiments were conducted to examine this hypothesis.

EXPERIMENT I

Method

General.—The first stage consisted of a training procedure designed to establish strong associations between the color of a target and its physical size. The intention was to train *S* to assume the size of the target from observation of its color. In the second stage, pairs of equidistant colored targets were presented under reduced conditions of observation, and *S* reported their relative distance. The size and color of the targets were systematically varied in a manner designed to reveal the effect of the earlier training on the judgments of distance.

Apparatus.—A modified version of the table used in earlier experiments (Epstein, 1961, 1963) was the main apparatus. This apparatus allows the standard and comparison objects to be presented in separate fields. The standard field can be viewed only monocularly, while simultaneously the comparison field is viewed binocularly. The standard alley contained two light boxes whose centers were 28 cm. above the table top (approximate eye level) and separated laterally by 22.5 cm. forming an angle of 10° with *S* as the vertex. The faces of the boxes were in *S*'s frontal plane at a distance of 150 cm. from the viewing aperture. The standard squares were inserted in holders in the faces of the boxes. Under the experimental conditions, only the standard squares were visible in the standard alley. The comparison alley was well lighted and contained two tracks which formed an angle of 10° with *S* as the vertex. Riding on each track was a cart which carried an ir-

---

[1] This work was supported by Research Grant MH 04153-03 made to the first author by the National Institute of Mental Health of the United States Public Health Service.

TABLE 1

SIZE-COLOR CONJUNCTIONS LEARNED BY THE EXPERIMENTAL *S*s

Color of Training Square	Size (In.) of Square during Training		
	Subgroup 1	Subgroup 2	Subgroup 3
Blue, yellow	4	2	1
Red, brown	2	1	4
Orange, green	1	4	2

regularly contoured nonsense form. The two forms were equal in area but were differently shaped. By pulling two strings *S* was able to vary the distance of the comparison forms. During the experiment *S*'s head was restrained in a chin-rest head-clamp arrangement to discourage movement.

Materials.—The standards were 1-, 2-, and 4-in. squares of layers of colored tissue paper mounted in black cardboard frames. Six colors were used: yellow, blue, brown, red, orange, and green.

Subjects.—Forty-two students from an introductory psychology course served as *S*s, 30 of whom were in the experimental group while 12 were control *S*s. The assignment to conditions was in a random order.

Procedure.—In the first stage the experimental *S*s performed an oddity task. On each of 48 trials a set of three squares differing in size and/or color was presented, and *S* was instructed to point to the square "which does not belong with the others." The conditions of the task were contrived to encourage *S* to learn the prescribed size-color conjunctions. Objective evidence that *S* learned these conjunctions was provided by a test of free recall following the last trial. There were three subgroups of 10 *S*s. Each group learned a different set of color-size conjunctions. The plan is shown in Table 1. By combining the data for these subgroups any effects of color per se on apparent distance (Johns & Summer, 1948; Over, 1962) would be distributed evenly among the various comparisons.

Immediately following the training task *S* began the second stage. This consisted of a series of judgments of the relative distances of the members of 20 pairs of squares. The 20 pairings can be classified into three categories: (1) Nine trials on which both members of the pair were from the same *trained* size category but differed in physical size when presented. (2) Six trials on which members of the pair were from different trained size categories but were the same physical size when presented. (3) Five trials on which the members of the pair were from different size categories and also differed in physical size when presented. Table 2 contains a schematic description of illustrative trials in each classification.

The *S* was instructed to report his immediate impression of the relative distance of the squares. The main part of the instructions was as follows:

> I will designate a square by calling out its color and I would like you to move the cart which is on the same side as the square (left or right) to a position along the track which is the same distance from you as the square. Then I will ask you to set the second nonsense form so that it is at the same distance in relation to the first form as is the second colored square to the first colored square.

Each *S* made one ascending and one descending judgment of the distance of each of the members of each pair. On half the trials *S* matched the right-hand square first and on the other trials he matched the left-hand square first. The position (left or right) of the members of each pair was varied between *S*s. The 20 pairings were presented in random order.

At the conclusion of the distance judgments each experimental *S* was tested for the retention of the color-size associations which he learned in the preliminary training. The *S* was shown a series of six black cardboard squares which duplicated the dimensions of the standards, and he was asked to designate the color which was associated with each size during the training.

The procedure for the control *S*s was the same except that the first stage was omitted.

Results

Inspection of the data did not reveal any systematic effect of direction of judgment (ascending or descending) or order of judgment (left or right square matched first) on the apparent radial separation of the standards. Therefore, the results for these conditions were combined. Table 3 shows the mean apparent distances of the standards for the experimental and control *S*s on each of the eight trials shown in Table 2. The results for the

TABLE 2

Schematic Description of a Sample of Trials and the Theoretical Expectations

Class	Left Square		Right Square		Predictions[b]	
	Size Trained[a]	Size Presented	Size Trained	Size Presented	Assumptional	Visual Angle
1	L	L	L	I	RF	RF
1	I	S	I	L	RN	RN
2	L	L	I	L	RN	E
2	L	I	I	I	RN	E
2	L	S	I	S	RN	E
3	L	L	I	I	E	RF
3	S	S	I	I	E	RN
3	L	L	S	I	RN	LN

[a] L = large, I = intermediate, S = small.
[b] RF = right square farther; RN = right square nearer; LN = left square nearer; E = squares equidistant.

remaining 12 trials are not given.[2] In every case the results for the trials which are omitted were consistent with the data for the reported trials in their respective classes.

The predictions in the two right-hand columns of Table 2 have been labeled assumptional and visual angle. The assumptional predictions are those derived from the theoretical account of relative size described in the introduction. The visual-angle predictions are made entirely on the basis of relative visual angle, disregarding the presumed influence of assumptions. Introducing assumptional considerations on Class 1 trials does not lead to predictions which differ from those based on visual-angle relationships alone. However, on the trials in Classes 2 and 3 the assumptional hypothesis leads to predictions which differ from those derived solely by taking into account the difference in visual angle. A comparison between the expectations presented in Table 2 and the data in Table 3 shows that in every instance the outcome was predictable entirely on the basis of relative visual angle.

TABLE 3

Mean Apparent Distances of Standards for the Trials Described in Table 2

Class	Experimental *Ss* (*N* = 30)			Control *Ss* (*N* = 12)		
	Left Square	Right Square	Apparent Radial Separation	Left Square	Right Square	Apparent Radial Separation
1	133	153	20	114	161	47
1	185	128	57	197	127	70
2	136	137	1	122	128	6
2	154	153	1	156	156	—
2	183	183	—	183	189	6
3	131	159	28	116	159	43
3	185	157	28	189	149	40
3	130	156	26	122	160	38

---

[2] Extended versions of Tables 2–5 which include the data for all the trials have been deposited with the American Documentation Institute. Order Document No. 7861 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

Table 3 also shows that the outcomes for the experimental and control *S*s were similar. The only effect of the pretraining seems to have been a reduction in the magnitude of the relative size effect for the experimental *S*s. The mean difference in the perceived distance of the standards was smaller for the experimental *S*s on 17 of the 20 trials. The difference between the overall mean separations for the 20 trials for the two groups was significant, $t\ (18) = 5.64$, $p < .01$. The diminution of the difference in perceived distance cannot be construed as evidence of the influence of assumptions. In fact on the trials in Class 2 the diminution contradicts the assumptional expectations. In general, the smaller separations which were obtained after the pretraining were not related to the assumptional thesis in any systematic manner.

The test of retention of the trained size-color conjunctions yielded unequivocal results. All *S*s exhibited complete retention.

EXPERIMENT II

Experiment I tried to establish specific assumptions experimentally by training *S* with the standards. It is possible, however, that this procedure failed to achieve its purpose. The training period was relatively brief, and may have been insufficient for the purpose of establishing a stable, compelling color-size entailment. It is true that *S*s exhibited perfect retention of the size-color conjunctions; however, the possibility remains that these conjunctions were not sufficiently overlearned to become spontaneously operative during the distance judgments. In addition, the color of an object is not normally a reliable prognostic indicator of size. This may have vitiated the effects of the training procedure. For these reasons an experiment was conducted to examine the influence of assumptions which *S* may be presumed to have developed through extended and frequent everyday commerce with the standards.

Method

Apparatus.—The apparatus was the same as in Exp. I except that the light boxes used to display the standards were removed. In Exp. II the standards were mounted at a distance of 135 cm. and illuminated by two focused beams of light which were restricted to the surfaces of the standards.

Materials.—There were nine standards, each a photograph of a dime, quarter, or half-dollar. Each coin was represented in its normal size, and also in the sizes of the two other coins; e.g., the quarter was reduced to the size of the dime and enlarged to the size of the half-dollar. The diameters of a normal dime, quarter, and half-dollar are 1.7, 2.5, and 3.0 cm., respectively.

Subjects.—Thirty-two new students from the same undergraduate population which served in Exp. I were the *S*s.

Procedure.—Relative-distance judgments of 23 pairs of standards were obtained following the procedure of Exp. I. The pairings may be classified in the same way as in Exp. I. There were 9 pairings in Class 1, 9 in Class 2, and 5 in Class 3. A sample of the pairings is represented in Table 4. The letters H, Q, and D represent the half-dollar, quarter, and dime photographs, respectively. The three sizes, large (normal, half-dollar size), intermediate (normal, quarter size) and small (normal, dime size) are represented by the letters L, I, and S, respectively. The 23 pairings were presented to each *S* in one of four different random orders. In addition, the following three variables were counterbalanced: temporal order of ascending and descending judgments; lateral position of the standard judged first; and the lateral position of each standard.

Results

Table 4 records the two predictions for each trial. The pairings in Classes 2 and 3 are of special interest since the introduction of assumptional considerations leads to expectations which differ from those dictated by visual-angle relationships. As an example Pairing 3c in Table 4 may be examined. In terms of relative visual angle HL should appear closer. However, if *S* assumes that the relative physical sizes of H and D correspond to their real-life relative sizes, then the larger than usual apparent size of D relative to H will cause DI to appear nearer than HL. For convenience, the data in Table 4 have been recorded in two columns labeled Left and Right. The means in the Left column represent the apparent distances of the left-hand members of the pairs listed in the Pairing column and the means in the Right column represent the apparent distances of the right-hand members. Thus for Pairing 1a DS appeared at 176.71 cm. and DI at 125.62 cm. Since the position of each coin was varied between *S*s, the designations left and right in Table 4 do not refer to the location of the standard.

The results shown in Table 4 are as unequivocal as those obtained in Exp. I. However, whereas Exp. I failed to provide evidence for the influence of assumptions, the results for all seven critical trials in Classes 2 and 3 of Table 4 support the assumptional hypothesis. This was also true for the remaining seven trials in Classes 2 and 3.

A supplementary analysis provided additional support for the assumptional interpretation. If the difference in the perceived distance between standards is mediated by assumptive factors, then the magnitude of the difference should be determined similarly. As an illustration, Pairing 2b may be compared with Pairing 2a (see Table 4). In both cases the assumptional hypothesis predicts separation while visual-angle considerations suggest equidistance. However, on closer consideration it will be apparent that the assumptional view also demands that the difference in perceived distance for Pairing 2b be greater than for Pairing 2a. This follows, in the most general way, from the fact that the discrepancy between assumed and actual physical sizes is greater fo Pairing 2b. Table 5 shows a sampl of the 12 comparisons which wer made. In order to make these comparisons more convincing we hav compared trials which presented th *same visual-angle relationships.* Tab 5 shows that the results for all th comparisons confirmed the requir

TABLE 4

SAMPLE OF PAIRINGS PRESENTED IN EXP. II: THEORETICALLY REQUIRED AND OBTAINED RELATIVE DISTANCE

Class	Pairing	Predictions		Obtained Apparent Distance		*t*
		Assumptional	Visual Angle	Left	Right	
1a	DS vs. DI	DI nearer	same	176.21	125.62	10.56*
1b	DI vs. DL	DL nearer	same	148.18	119.56	7.22*
2a	DS vs. QS	D nearer	equidistant	150.53	193.84	7.95*
2b	DS vs. HS	D nearer	equidistant	151.12	213.34	9.48*
2c	DI vs. QI	D nearer	equidistant	127.96	158.15	5.90*
2d	DI vs. HI	D nearer	equidistant	130.59	177.40	6.14*
3a	QI vs. DS	equidistant	Q nearer	147.71	154.03	0.35
3b	HL vs. QI	equidistant	H nearer	152.43	157.68	1.18
3c	HL vs. DI	D nearer	H nearer	159.56	133.21	3.39

Note.—Letter subscripts in class designations are used to distinguish pairings within the same class.
** $p < .01$.

TABLE 5

PREDICTED AND OBTAINED APPARENT RADIAL SEPARATIONS FOR COMPARISONS BETWEEN PAIRINGS

Predicted Radial Separations for Selected Comparisons	Obtained Radial Separations			
	Pairing	Separation	Pairing	Separation
1a > 3a	1a	51.09	3a	6.32
1b > 3b	1b	28.62	3b	5.25
2b > 2a	2b	62.22	2a	43.31
2b > 2c	2b	62.22	2c	30.19
2d > 2c	2d	46.81	2c	30.19

ments of the assumptional hypothesis. The same was true for the remaining seven comparisons.

Especially convincing are two special cases not included in Table 5. These are comparisons between Pairing QI vs. QL and Pairing DI vs. HL, and between Pairing HS vs. HI and Pairing DS vs. HI. In these cases the assumptional hypothesis predicts a difference in the *direction* of apparent separation despite the identical retinal stimulation. The standards subtending the *larger* visual angles should appear nearer when both the members of the pair have the same identity. In contrast the standard subtending the *smaller* visual angle should appear nearer when the members of the pair are different coins. These expectations were confirmed.

DISCUSSION

The two experiments produced conflicting results. It might be argued that the pretraining task in Exp. I failed to modify the long established assumptions which *S* brought with him into the experimental situation. Nevertheless, in so far as the present findings are concerned, any positive conclusions about the role of assumptions must be restricted to the case of representative objects of known size. In this connection it is important to note that the relative perceived distance cannot be predicted by treating each standard as if it had been presented alone. If the formula derived by Epstein (1963, p. 261) from the known size-apparent distance hypothesis (Epstein, Park, & Casey, 1961, pp. 503–507) is applied to each member of the standard pair the outcome does not agree with the judged distances obtained in the present experiment. This occurs despite the fact that the formula provides remarkably accurate predictions for the case when a single standard is presented (Epstein, 1963). Evidently the simultaneous presence of two standards affects the perceived distance of each.

If the pretraining in Exp. I actually did modify *S*'s assumptions about size identity, then there remains the original problem of accounting for the consistent unidirectional results which are usually obtained with nonrepresentational standards. Alternative resolutions are not hard to formulate; however, they do not appear promising. As an example, consider the *minimum principle* (Hochberg, 1957) which has been successfully applied to other instances of depth perception in which *S* shows a preference for one of several alternative perceptual outcomes all of which are equivalent in terms of the retinal stimulation (e.g., Hochberg & Brooks, 1960). This is the principle that the perceptual system decodes the stimulus imput in accordance with the requirements of informational economy. More specific to the present problem, the minimum principle implies that if several alternative percepts are possible, then the percept which requires the least amount of information to specify will have the highest probability of occurrence. In order to apply this principle a ranking is needed of the alternatives in terms of the information which is required to specify each one. It is difficult to see the basis for assigning ranks in the present experiments.

We should note, however, that the hypothesis of Ittelson and Ames is also not without difficulties. These authors assert that *S* must, on a conscious or subconscious level, make the assumption of identity in order that the reported perceptual experience occur. This view depends ultimately for its validity on evidence of the functioning of assumptions *independent* of the perceptual response which is presumed to be the consequence of these assumptions. This evidence is not available. In the absence of such data all that one can say is that *S* responds *as if* he were making the assumption of identity. In this formulation the hypothesis is considered to be explanatory but not necessarily descriptive of the psychological processes which mediate the effect. Therefore, its validity will depend on successful prediction, rather than demonstrations of the reality status of the identity assumption.

REFERENCES

AMES, A., JR. *Some demonstrations concerned with the origin and nature of our sensations.* Hanover, N. H.: Institute for Associated Research, 1946.

EPSTEIN, W. The known size apparent distance hypothesis. *Amer. J. Psychol.*, 1961, **74,** 333–346.

EPSTEIN, W. The influence of assumed size on apparent distance. *Amer. J. Psychol.*, 1963, **76,** 257–265.

EPSTEIN, W., PARK, J., & CASEY, A. The current status of the size-distance hypotheses. *Psychol. Bull.*, 1961, **58,** 491–514.

HOCHBERG, J. Effect of the Gestalt revolution: The Cornell symposium on perception. *Psychol. Rev.*, 1957, **64,** 73–84.

HOCHBERG, J., & BROOKS, V. The psychophysics of form: Reversible perspective drawings of spatial objects. *Amer. J. Psychol.*, 1960, **73,** 337–354.

ITTELSON, W. H. *Visual space perception.* New York: Springer, 1960.

JOHNS, E. H., & SUMMER, F. C. Relation of the brightness differences of colors to their apparent distances. *J. Psychol.*, 1948, **26,** 25–29.

OVER, R. Stimulus wavelength variation and size and distance judgments. *Brit. J. Psychol.*, 1962, **53,** 141–148.

(Received June 21, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 514–522

CONCEPT LEARNING AND PROBABILITY MATCHING[1]

GEORGE MANDLER, PHILIP A. COWAN,[2] AND CECILLE GOLD[3]
University of Toronto

Ss learned 1 of 3 concepts varying in level of difficulty. For each concept $\frac{2}{3}$ of the instances were positive, $\frac{1}{3}$ negative. Ss matched the frequency of the response appropriate for the positive (negative) instances well in advance of having learned the concept. A group which was given prior training in such probability matching showed slightly improved concept performance; prior training with random information tended to retard concept learning. Reversal of the response requirement for positive and negative instances interfered with performance when Ss had learned the concept only to a low level; it did not interfere with mastery of a well-learned concept.

Mandler and Cowan (1962) have reported that in a situation where the ostensible task requires one particular response for a number of different stimulus events, *S*s will, on the average, emit that response at the required frequency. Such probability matching occurs even when *S*s have not yet solved the task. These data suggested two parallel learning processes; *S*s learn how frequently a particular response is required (probability matching) and they also learn *when* that response is required (event matching). In the Mandler and Cowan study it appeared that probability matching not only preceded event matching, but also that probability matching was more likely to occur the more difficult the ostensible task was.

The present study will extend that finding by examining trial-by-trial learning of concepts—an aspect of concept learning that has received relatively little attention (cf. Hunt, 1962). The distinction between probability matching and event matching will be explored by investigating the effects of different kinds of prior tasks involving probability matching on concept learning, and by comparing probability matching during the acquisition of concept tasks of varying difficulty. Finally, the effect on concept learning of response reversal, which leaves the nature of the concept unchanged, will be examined.

METHOD

Tasks.—The major requirement in constructing the concept tasks was that they should vary in difficulty, and that the response required should permit not only correct responding in line with the event information provided, but should also permit responding to the probability structure without attending to the event information.

Three different tasks were constructed: Easy, Medium, and Difficult. On each task *S* was given a set of 3 × 5 in. cards on the face of which were printed the instances of the concept and on the back of which were check marks for the positive instances and no check marks for the negative instances. The *S*s were required to look at the front of a card and mark on it with a pencil whether the back had a check mark or not. They then looked at the back and went on to the next card. A trial was defined as six successive cards containing four positive and two negative instances.

The same three dimensions were used for the three tasks. They were:

1. *Letters* of the alphabet: A B C D E F
2. *Number* of letters: 1 2 3 4 5 6
3. *Type:* Upper and lower case.

---

[1] This research was supported in part by Grant GB-810 from the National Science Foundation, and by Grant APA-37 from the National Research Council, Canada.

[2] Now at University of California, Berkeley.

[3] Now at Goucher College.

For the Easy task all three dimensions were relevant and redundant. Four of the *letters*, four *numbers* of letters, and one *type* always were positive instances and the negative instances were the other two letters, two numbers, and type. In addition the same letters were paired with the same numbers. For example, for a particular *S* the following positive instances appeared on each trial: 5 upper A's, 1 upper C, 6 upper Ds, 4 upper Fs; and negative instances: 3 lower Bs, 2 lower Es. In order to learn the concept, *S* could identify the four relevant letters, or the four relevant numbers, or the one relevant type. Twenty such "trials" were randomly selected and used for the 20 *Ss* in each subgroup who were given the Easy task.

For the Medium task, only letters and numbers were relevant and redundant. For this task a concept is defined in two trials (12 cards) with type varying across the two trials. Fifteen such sets of two trials (12 instances) were selected with five of them repeated to produce 20 concepts for 20 *Ss*.

It should be noted that for the Easy and Medium tasks the same instances were repeated in each trial or two trials, respectively. The order of instances within trials was randomized with the additional restriction that no instance ever occurred twice in succession.

For the Difficult task only one dimension was relevant, namely *number* of letters. This concept is defined by four of the six numbers always being positive instances and the other two negative, and 72 cards (12 trials) are needed to define the concept. Here again 15 combinations were selected, with 5 of them repeated to produce 20 concepts for 20 *Ss*. It should be noted that repetition of instances occurred only after 12 trials (72 cards).

During response reversal the stimulus cards remained the same but the response requirement of check marks for positive and no check marks for negative instances was reversed to check marks for negative and no check marks for positive instances.

Experimental design.—The design of the study is shown in Table 1. Within each of the three concept groups, independent groups of *Ss* received three kinds of prior experience. The Information (Inf) groups learned the concept without prior experience. They received 24 trials on the concept (Stages I and II) followed by 12 trials of response reversal. The Blank (Bl) groups were first given 12 trials (Stage I) with cards with blank fronts, but check marks on the back just as the regular Inf *Ss*. During this stage *Ss* had to predict check marks without any additional information, i.e., they were given prior probability matching. The Random (Ra) groups, during Stage I, received 12 trials with information on the front of the cards identical to that given to the Inf groups, but with check marks assigned to the cards at random, except that within each trial of six cards there were still four check marks and two non-check-marks. These *Ss* thus were given prior random experience. All Bl and Ra groups were switched to their respective concept tasks at the beginning of Stage II and following 12 such trials they were switched to 12 trials of response reversal in Stage III.

In addition to these nine groups (three

TABLE 1

EXPERIMENTAL DESIGN

Experimental Group	Concept Level	*N*	Concept Problem in Stages of 12 Trials (72 Cards)		
			Stage I	Stage II	Stage III
Information (Inf)	Easy Medium Difficult	20 20 20	Easy Medium Difficult	Easy Medium Difficult	Easy Reversal Medium Reversal Difficult Reversal
Blank (Bl)	Easy Medium Difficult	20 20 20	Blank Blank Blank	Easy Medium Difficult	Easy Reversal Medium Reversal Difficult Reversal
Random (Ra)	Easy Medium Difficult	20 20 20	Random Random Random	Easy Medium Difficult	Easy Reversal Medium Reversal Difficult Reversal
Information Control (Inf C)	Easy Medium Difficult	20 20 20	Easy Medium Difficult	— — —	— — —

concepts with three types of prior activity) of 20 *S*s each, three additional groups were run—the Information Control (Inf C) groups. These *S*s were given one of the three concept tasks just as the Inf groups, except that the responses required for positive instances (check marks) were used for negative instances, and those required for negative instances (non-check-marks) were used for positive instances. These three groups of *S*s were given the 12 trials for Stage I only.

Subjects.—The *S*s were 240 students taking introductory psychology courses at the University of Toronto. Seventy-five per cent were female and 25% male, and this proportion was maintained in all experimental groups. Twenty *S*s were assigned at random to each of the 12 experimental groups.

Procedure.—The *S*s were run in groups of 60, i.e., three experimental groups at a time. Each *S* was given a pack of 216 cards (except for *S*s in the Inf C groups who only received 72 cards). The instructions were essentially the same as those used by Mandler and Cowan (1962). The operative part instructed *S*s that they were to try to predict whether the back of a card had a check mark on it or not. They were told that some cards would have check marks and some not, that some of the cards might have material printed on the front and some not, and that they were to pay attention to the material on the front which might help them to learn where to predict check marks. The *S*s' task was to place a check mark on the front of the card if they believed that there was a check on the back and to place a line on the front if they believed there was no check on the back. They were paced at a rate of 7 sec. per card by *E* telling them when to turn over a card, look at the back, look at the next one, predict, and so forth. Since all 216 cards were in one deck, *S*s were given no warning when the task changed from one stage to another.

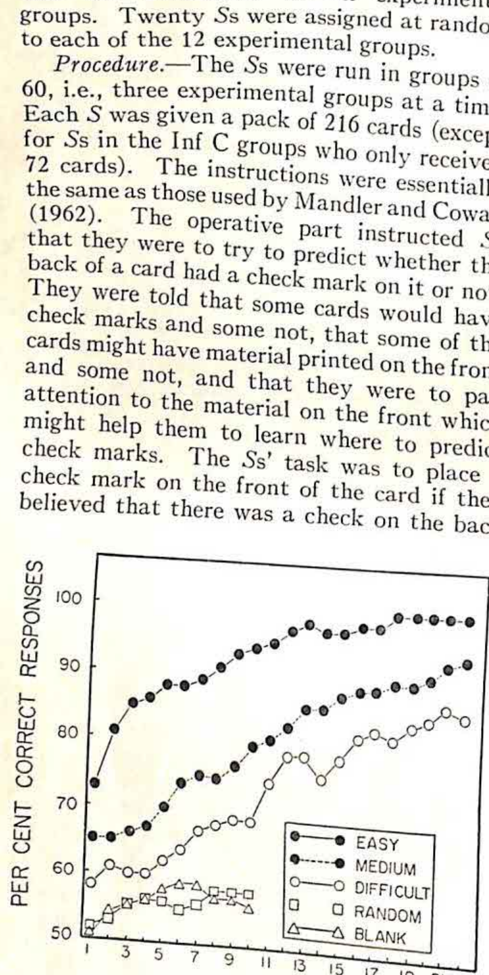


FIG. 1. Frequency of correct responses for 24 trials of the Inf groups on three concepts (Easy, Medium, and Difficult) and for the first 12 trials of the Ra (Random) and Bl (Blank) groups.

RESULTS

Concept learning.—The three upper curves in Fig. 1 show concept learning for the 24 trials of Stages I and II for the three Inf groups (Inf Easy, Inf Medium, and Inf Difficult). The ordering of the three learning curves is as expected. An overall analysis of variance was performed on the first 12 trials of concept learning for the four experimental groups (Inf, Bl, Ra, and Inf C) under three concept conditions. This analysis compares Stage I on Inf and Inf C with Stage II of Bl and Ra. While the difference among concepts (Easy, Medium, and Difficult) was highly significant, $F\ (2,\ 228) = 44.6,\ p < .001$, the four groups do not differ significantly, $F\ (3,\ 228) = 1.74,\ p < .25$, nor is the interaction between groups and concepts significant. The effect of prior experience will be examined in detail below.

The two bottom curves in Fig. 1 show the performance during Stage I of the Bl and Ra groups, with the three concept groups combined in this case. During this stage, i.e., prior to concept learning, Bl and Ra are essentially identical, but both groups show a significant learning effect, $F\ (11,\ 1254) = 2.56,\ p < .01$.

Probability matching during concept learning.—The upper curves of Fig. 2 show the frequency of the check-mark responses (probability matching) for the Inf Easy, Medium, and Difficult groups during Stages I and II as well as for the Bl and Ra groups during Stage I. It is clear that all three concept groups reach asymptote before

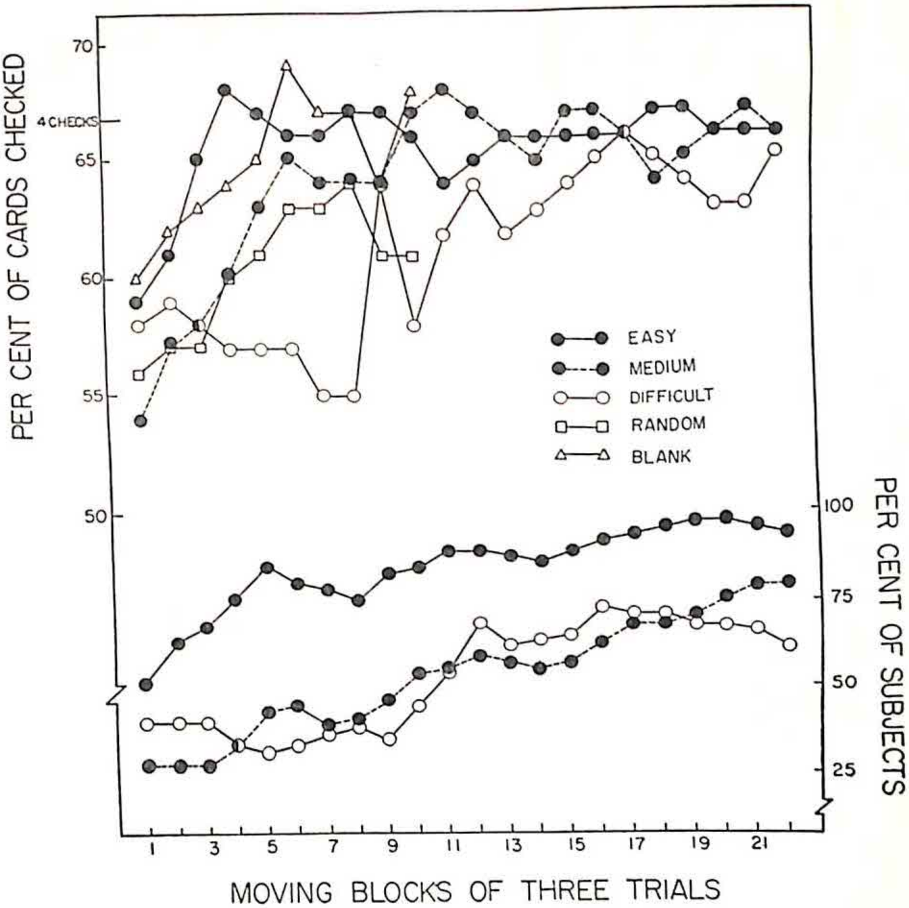


FIG. 2. Frequency of check-mark responses (probability matching) for 24 trials (Stages I and II) of the Inf groups on three concepts (Easy, Medium, and Difficult) and for 12 trials (Stage I) of the Ra (Random) and Bl (Blank) groups (upper curves); and percentage of *S*s in the three Inf groups giving perfect probability matching (four check marks) per trial (lower curves).

perfect event performance is reached (cf. Fig. 1). The Inf Easy group reaches 67% by Block 4, the Inf Medium group by Block 10, and the Inf Difficult group reaches 66% by Block 17. The difference among these three groups is significant, $F\ (2, 57) = 5.53, p < .05$.

It could be argued that these average levels of frequency matching need not necessarily reflect probability matching on the part of individual *S*s. The lower graph in Fig. 2 addresses itself to this question. It shows the percentage of *S*s, in each of the three Inf concept groups during Stages I and II, producing four check marks and two non-check-marks on any given block of trials. The data show that 50% of the *S*s in the Easy concept group match check-mark frequencies during the first block of trials, reaching close to 100% by the last blocks. For the Medium and Difficult concepts the increase is from about 25% to about 75%. The question to be raised is whether more *S*s per trial match the probability structure than match the event structure. Obviously, in the Easy concept the level of probability matching is high because the level of event matching is high (cf. Fig. 1) and when *S*s have solved the task, they match the frequency of check marks by definition. On the other hand, during the last 12 of the Medium and Difficult concept trials only 56.6% of the 480 *S*-trial units (12 trials for each of 40 *S*s) show perfect event matching, while 67.0% show perfect probability matching. As a statistical test of this finding, each of the last 12 trials for the Medium and Difficult concepts was given one score for the number of *S*s (out of 20) who gave completely correct event-matching responses and another score for the number of *S*s who matched the probability structure perfectly. An analysis of variance showed a significant finding for the difference between event matching ($M = 11.3$ *S*s per trial) and probability matching ($M = 13.4$ *S*s per trial), $F\ (1, 23) = 41.93$, $p < .0001$. Since *S*s may, however, assign four check marks in any one trial by chance, the frequency of this event in 224 trials contributed by *S*s who had not yet mastered the concept was compared with the chance expectancy derived from the binomial distribution. The data showed that probability matching occurred in 32.6% of these trials, compared with a chance expectancy of 23.4%, $\chi^2\ (1) = 8.05$, $p < .01$.

As far as the level of probability matching for Stage I of the Ra and Bl groups is concerned, the level of check-mark production of the Ra group is consistently and significantly below that of the Bl group, $F\ (1, 114) = 8.90$, $p < .005$.

Mandler and Cowan (1962) had found that probability matching was more quickly attained the more difficult the task was. In that study, however, difficulty of tasks was directly related to the nature of the more frequent response. When the identical task was used in two groups, but one group used a check mark as the more frequent response while the other used non-check-mark as the more frequent response, it was found that the group using check marks learned the correct response better and showed lower response frequency, i.e., they did not match the probability structure. This held for two comparisons, when the more frequent response occurred 67% or 83%. In the present study, the three Inf C groups were added specifically to test this finding. It will be recalled that these three groups are identical with the Inf groups except that the response requirement is changed from 67% check marks to 67% non-check-marks. A comparison of these three sets of groups showed a failure to replicate the finding of the previous study. The comparable Inf and Inf C groups did not differ consistently or significantly with respect to either probability matching or concept learning.

Prior experience and concept learning.—Figure 3 shows, for each of the three concepts, 12 trials of concept learning of Stage I of the Inf groups and Stage II of the Bl and Ra groups. The overall ranking of the three experimental groups is the same on all concepts—Bl first, then Inf, followed by Ra. The probability that such an ordering obtained with one set of three independent groups could be replicated by chance by another two sets is 1/36 ($p = .03$). While the overall analysis of variance failed to show any significant effects for groups, the Bl groups are significantly better than the Ra groups, $F\ (1, 114) = 4.36$, $p < .05$. In terms of number of *S*s reaching criterion of two successive perfect trials the data show superior performance of the Bl *S*s, 53.3% of whom reach criterion as against 36.7%

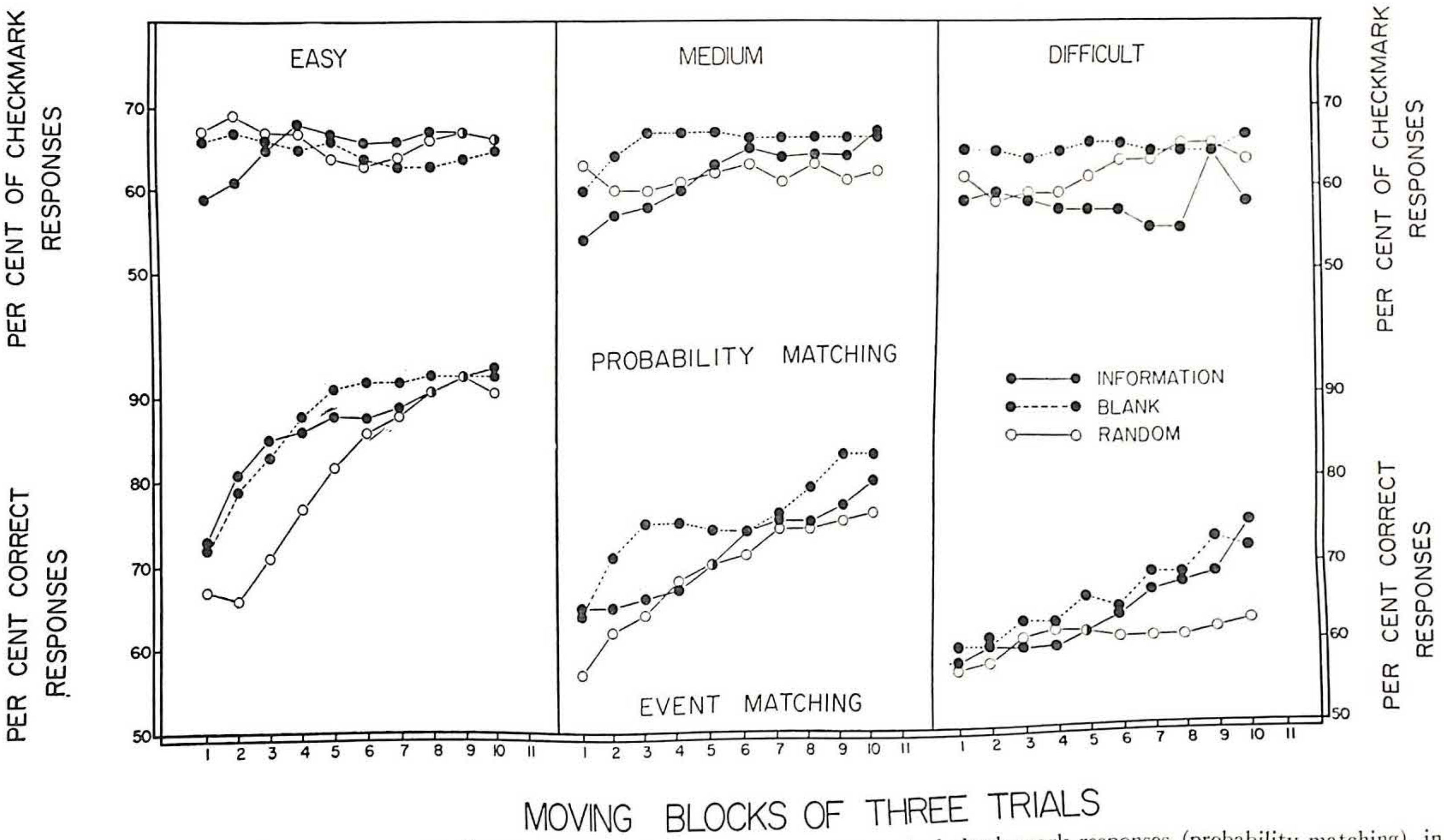


FIG. 3. Frequency of correct responses (event matching), in lower three curves, and of check-mark responses (probability matching), in upper three curves, for each of three concepts on the first 12 trials (Stage I) for the Inf groups and Trials 13–24 (Stage II) for the BI andRa groups.

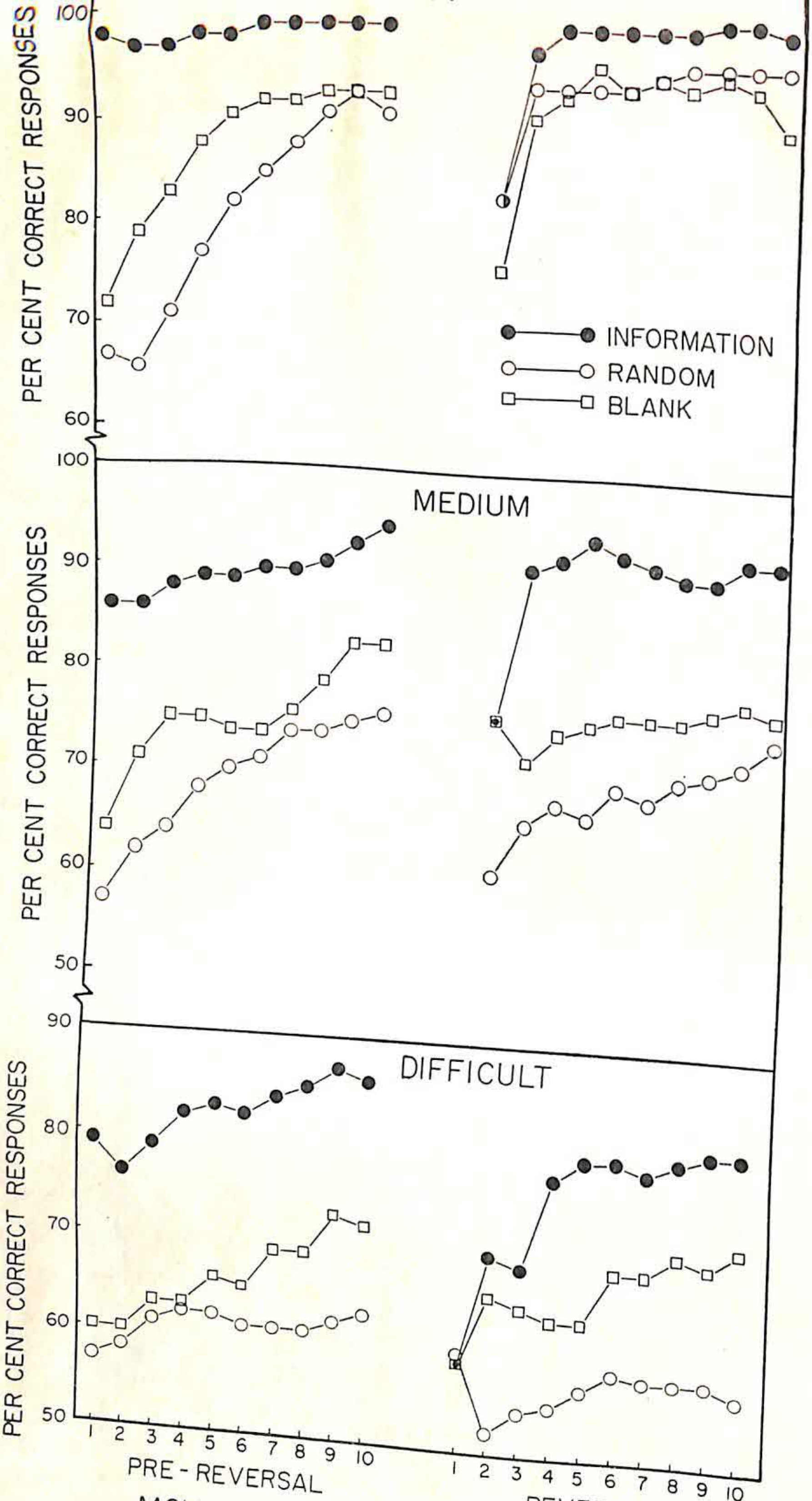


FIG. 4. Per cent correct performance on each of three concepts for the Inf, Bl, and Ra groups immediately prior to (prereversal) and following reversal of the response requirement.

of the Ra and Inf *S*s, $\chi^2\ (1) = 4.60$, $p < .05$. There were no interactions between these groups and difficulty of the concept.

Inspection of the probability-matching curves in Fig. 3 shows that the Bl groups maintain a high level of asymptotic responding, which is not the case for the Inf and Ra groups on the Medium and Difficult concepts. The high level of probability matching on the Easy concept is, of course, at least in part forced by the high level of event matching for this concept.

Response reversal.—Figure 4 shows the performance of all nine experimental groups (three concepts and three prior learning conditions) during Stage II (prereversal) and Stage III (reversal). For the Easy concept there is a very rapid recovery to prereversal levels upon change of the response requirement. For the Medium and Difficult concepts, however, all three groups show a decrement in performance following response reversal. Performance on the last two trials of Stage II (prereversal) was compared with performance on Trials 2–12 of Stage III. Trial 1 of Stage III was eliminated since on this trial *S*s did not yet know that the response had been reversed and would necessarily make a large number of errors. While there were no differences among the Inf, Bl, and Ra groups as a whole, the Easy concept showed that only 42% of the *S*s performed worse following reversal, while 56% of the *S*s on the Medium and Difficult concepts performed worse following reversal, $\chi^2\ (1) = 8.45$, $p < .01$.

DISCUSSION

The data show that *S*s will learn the frequency of a response prior to learning correct event matching, i.e., when and where that response is to be made, and thus bear out the previous findings of Mandler and Cowan (1962). However, contrary to a suggestion contained in that study, level of probability responding varied directly with task difficulty, such that the more difficult the task the lower the level of the frequency response and the later asymptote is reached.

Frequency matching without event information (the Bl group) was about at the same level as that exhibited during performance of the Easy task (Inf Easy). However, the more difficult the event information the lower the level of the frequency response. This deterioration of probability matching with the addition of event information is also suggested by the behavior of the Ra group. Here random event information resulted in a frequency response significantly below that of the Bl group. What might be the case is that the kinds of hypotheses that *S*s developed on the basis of the event information frequently diverge from the probability structure of the task and thus lower matching behavior. If we assume that asymptotic probability matching is ascribable to *S*s' uses of sequential hypotheses that reflect the frequency of the input, then the addition of *concept* hypotheses might result in some deviation from asymptotic levels.

Prior experience with probability matching apparently had a slight effect in improving subsequent concept attainment. The effect of prior training with a random task apparently was a lowering of performance on the concept task. The most obvious explanation would suggest that *S*s' failures to find a satisfactory hypothesis during the random trials is likely to interfere with their attainment of the correct hypothesis, i.e., the concept.

Reversal of the response requirement resulted in deterioration of concept performance on Medium and Difficult concepts. In other words, at a low level of concept attainment, response reversal affects conceptual behavior. When, as in the Easy concept, the concept rule has been well learned, response reversal has practically no deleterious effect on performance. This finding suggests that the mastery of the mediating rule determines in part how well reversal shifts can be accomplished. When the conceptual rule is well established, a reversal of previously learned responses does not interfere with its application.

The data on concept learning and probability matching indicate that these two processes can be separated and that they do go on concurrently. It should be noted that matching probabilities results in some improvement in event-matching performance. How much such improvement occurs is indicated by the quasi-learning curves for Stage I of the Bl and Ra groups (Fig. 1) where correct predictions do increase as the probability-matching curve approaches asymptote.

Either probability matching or event matching will yield better than chance success. When no other alternative is available to *S*s (as in the Bl and Ra groups) probability matching alone will occur and produce improved performance. When both response principles are available, however, *S*s apparently will match the probability structure as well as engage in the more efficient event-matching behavior. Thus it was shown that, *prior* to mastery of the two more difficult concepts, probability matching occurs more frequently than event matching and more frequently than would be expected by random production of the check-mark responses.

REFERENCES

HUNT, E. B. *Concept learning.* New York: Wiley, 1962.

MANDLER, G., & COWAN, P. A. Learning of simple structures. *J. exp. Psychol.*, 1962, **64**, 177–183.

(Received June 24, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 523–530

UNCERTAINTY, INFERENCE DIFFICULTY, AND PROBABILITY LEARNING[1]

CAMERON PETERSON AND Z. J. ULEHLA[2]

University of Colorado

On the basis of information provided by a cue, Ss inferred which of a set of criteria would occur. 2 different problems were studied using this experimental task. Problem I evaluated the effects of 4 different information-theory variables upon 5 different measures of inference difficulty. Difficulty of inference increased with increased uncertainty of the criteria not resolved by the occurrence of the cue. Problem II studied the relations between 3 different conditional probabilities associated with the most frequently occurring criterion. Response probability approximated or exceeded the corresponding probability of occurrence, which in turn exceeded the corresponding subjective probability.

A single experiment investigated two different problems involving the relation between quantitative measures of the stimulus situation and dependent performance. These problems are treated separately in both the introduction and the results sections.

Problem I

An important application of information theory to psychology is its use as a mathematical model for predicting difficulty of stimulus identification. The general hypothesis is that, as the uncertainty of a stimulus increases and thus as the amount of information-processing activity required to identify the stimulus increases, performance in the identification of a stimulus deteriorates. Supporting this hypothesis are the findings that: latency from stimulus presentation to stimulus identification is generally a linear function of the uncertainty of the stimulus (Brainard, Irby, Fitts, & Allusi, 1962); accuracy of pattern reproduction and pattern identification decreases as pattern uncertainty increases (Attneave, 1955); and language reproduction decreases as the uncertainty of the words increases (Miller & Selfridge, 1950). Taken together, these various types of performance decrement support the general hypothesis that stimulus-identification difficulty increases with stimulus uncertainty.

However, much information processing goes beyond stimulus identification to the use of the stimulus in making inferences about future events. Such tasks may be termed inference tasks. The basic paradigm for an inferential task is that an *S* observes the occurrence of a cue and, on the basis of the information provided by the cue, predicts which of a set of criteria will occur. Just as the amount of uncertainty associated with a stimulus determines the difficulty involved in identifying it, so may the uncertainty associated with the inference task determine inferential difficulty.

---

[1] The research reported here was undertaken in the Behavior Research Laboratory, Institute of Behavioral Science, University of Colorado, and is Publication No. 35 of the Institute. This research was supported by a Research Grant (M-4977) from the National Institute of Mental Health. Grateful acknowledgment is made by the authors to K. R. Hammond and F. J. Todd for valuable suggestions and to D. Derrer and R. Lehman for their assistance in the conduct of this experiment.

[2] Now at the University of Denver.

Problem I concerned an extension of the uncertainty-difficulty hypothesis from stimulus identification to inferential tasks.

In a typical stimulus-identification task, information is provided from the environment in a single step at the time that the stimulus event occurs. In an inference task as described above, the information is provided in two steps: (*a*) when the cue occurs, and (*b*) when the criterion occurs. Thus there are two forms of uncertainty in an inference task, one concerns which cue will occur and the other concerns which criterion will occur. As in the stimulus-identification task, cue uncertainty, $H(\text{Cue})$, is completely resolved upon occurrence of a cue. Criterion uncertainty, $H(\text{Criterion})$, on the other hand, may be resolved in connection with both of the steps. If cues provide information about which criteria will occur, part of $H(\text{Criterion})$ may be resolved upon occurrence of a cue; the remainder cannot be resolved until the occurrence of the criterion. The portion of $H(\text{Criterion})$ resolved upon occurrence of the cue is defined as information transmitted from cue to criterion, $T(\text{Cue; Criterion})$; the remaining uncertainty is defined as conditional criterion uncertainty, $H(\text{Criterion}|\text{Cue})$. Thus, there are several interrelated but distinguishable uncertainty measures relevant to an inference task: (*a*) $H(\text{Cue})$, (*b*) $H(\text{Criterion})$ and its components $T(\text{Cue; Criterion})$ and $H(\text{Criterion}|\text{Cue})$. In pursuit of Problem I, all of these measures of uncertainty were manipulated in order to ascertain their relation to inference difficulty.

In order to ascertain this relationship, it was necessary to measure inference difficulty. Several difficulty measures were employed in an effort to triangulate on the construct of inference difficulty. Some of the measures, as in the case of stimulus identification, focus on performance decrements; other measures involve evaluation of difficulty by *S*s. Specifically, five measures of difficulty were used in this experiment: (*a*) proportion of predictions of the event having the highest frequency of occurrence (maximizing), (*b*) proportion of predictions which turn out to be correct (hit rate), (*c*) reaction time, (*d*) *S*s' rankings of difficulty, and (*e*) *S*s' confidence in the accuracy of their predictions (confidence). These difficulty measures combine apparent face validity with operational divergence. This experiment tested the effect of each of the four uncertainty measures upon all five measures of difficulty.

Problem II

Orderly relations have been observed between certain probabilities associated with the most frequent of a set of alternative events: π, the relative frequency of occurrence; $P(R)$, proportion of trials on which *S* predicts the most frequent event; and $P(S)$, *S*s' estimates of the probability of the most frequent event. In very simple probability-learning situations, asymptotic $P(R)$ has repeatedly approximated or slightly exceeded π (Edwards, 1961). In other experimental situations, $P(S)$ proved to be underestimates of π (e.g., Attneave, 1953). If these two relations were to characterize the same experimental situation, then a third relation is implied—$P(R)$ must exceed $P(S)$. Along with investigating Problem I, this research sought, as Problem II, to determine if these three relations would characterize several levels of π in a complex learning situation. As π varies, will $P(R)$ approximate, perhaps slightly exceed, π, and will $P(S)$ be exceeded by both?

METHOD

A single experiment was designed to furnish measures relevant both to Problem I and to Problem II. In order to do this it was necessary to vary, as independent variables, the uncertainty measures mentioned above and, at the same time, to vary π. Dependent variables included the five difficulty measures involved in Problem I and $P(S)$ and $P(R)$ involved in Problem II.

Subjects

The 16 *S*s were paid university students.

Apparatus

The apparatus included numerous decks of cards, a device for eliciting normalized subjective probabilities, and a device for measuring reaction time.

The basic apparatus for a single learning trial consisted of a $1\frac{1}{2} \times 2$ in. card. The face of each card was either white, red, yellow, or blue, with color serving as the cue variable for all learning tasks. The four possible criteria for a given trial (card) were listed on the colored face, with a small box to the left of each criterion in which *S* marked his prediction with a pencil before viewing the correct criterion on the back of the card. The eight criterion variables used with their respective outcomes were Cars (Chevy, Ford, Plymouth, and Rambler), Names (George, Mike, Sam, and Ted), Letters (a, m, p, and t), Verbs (kick, run, strike, and throw), Nature (grass, lake, rock, and tree), Nonsense Syllables (dax, nef, yim, and zuk), Clothing (blouse, coat, hat, and shoe), and Animals (cow, deer, horse, and sheep).

The apparatus for obtaining normalized subjective probabilities consisted of a 25-in. rod marked in hundredths. Movable markers functioned to divide the rod into four different sections. Each of the four sections on the rod represented the probability of occurrence of one of the four criteria about which *S* made a probability estimate. The entire rod represented a probability of unity; thus the sum of the four mutually exclusive sections automatically summed to unity.

The apparatus for measuring reaction time enabled the cue side of a card to be presented to *S*. The presentation of the card started a timer which was stopped by a voice microphone when *S* verbally predicted which criterion was on the back of the card. Thus, reaction time was equivalent to latency between cue presentation and *S*'s prediction of the criterion.

Procedure

Experimental conditions.—Problem I and Problem II together required the manipulation of five independent variables: H(Cue), H(Criterion), T(Cue; Criterion), H (Criterion|Cue), and π. The combination of the first three variables in a factorial design, by appropriate variation in cue-criterion frequencies, resulted in the manipulation of all five variables. This $2 \times 2 \times 2$ factorial design created the eight different cue-criterion frequency matrices which served as the experimental conditions. These frequency matrices, along with the corresponding values of each independent variable, are shown in Table 1.

H(Cue) was varied by .19 bit; H(Criterion) and T(Cue; Criterion) were each varied by approximately .71 bit. Since H(Criterion|Cue) equals H(Criterion) minus T(Cue; Criterion), it was also varied, in each of two steps, by .71. The small variation in H(Cue) as compared with the other uncertainty variables was required in order to permit the unlikely cues in the low H(Cue) matrices to occur on an appreciable proportion ($\frac{1}{8}$) of the trials.

These variations of the joint cue-criterion frequencies made in conformance with the demands of Problem I also yielded the three indicated levels of π, thus providing additionally a manipulation of the independent variable of Problem II. In this experiment π will refer to the conditional probability of most frequent criterion, given the cue.

For each *S* the joint cue-criterion frequencies of each matrix in Table 1 were embodied in a deck of 320 cards. In order to minimize individual differences as a source of error variance, each *S* concurrently performed in experimental conditions represented by all eight matrices. In order to minimize transfer effects between conditions, a different one of the criterion variables described in the apparatus section was used for each condition. Any effect attributable to the differences among these criterion variables was minimized by counterbalancing the assignment of criterion variable to uncertainty conditions across *S*s. Each 320-card deck, embodying one condition for one *S*, was divided into four quarters, each retaining the prescribed cue-criterion probabilities as closely as possible. Within each quarter, the decks were shuffled to induce a nonsystematic order of presentation. The first quarter furnished the cards for Blocks 1 and 2, the second quarter, Blocks 3 and 4, and so forth.

Response measures.—The procedure was divided into two stages: group sessions and individual sessions. All but two of the

TABLE 1

Relative Frequencies in Cue-Criterion Matrices Embodying the Independent Variables

	Low T(Cue; Criterion)								High T(Cue; Criterion)							
	Low H(Cue)				High H(Cue)				Low H(Cue)				High H(Cue)			
Low H(Criterion)	24	72	14	42	48	48	28	28	36	108	4	0	72	72	8	0
	2	0	2	0	4	0	4	0	4	0	0	0	8	0	0	0
	14	42	24	72	28	28	48	48	0	12	36	108	0	8	72	72
	0	6	0	6	0	4	0	4	0	0	0	12	0	0	0	8
H(Cue)	1.81				2.00				1.81				2.00			
H(Criterion)	1.28				1.29				1.28				1.29			
T(Cue; Criterion)	.09				.08				.80				.82			
H(Criterion | Cue)	1.19				1.21				.48				.47			
π	.60				.60				.90				.90			
High H(Criterion)	16	18	8	30	32	12	16	20	24	0	2	42	48	0	4	28
	10	48	6	24	20	32	12	16	14	72	0	6	28	48	0	4
	8	30	16	18	16	20	32	12	2	42	24	0	4	28	48	0
	6	24	10	48	12	16	20	32	0	6	14	72	0	4	28	48
H(Cue)	1.81				2.00				1.81				2.00			
H(Criterion)	1.99				2.00				1.98				2.00			
T(Cue; Criterion)	.09				.09				.80				.80			
H(Criterion | Cue)	1.90				1.91				1.19				1.20			
π	.40				.40				.60				.60			

response measures were obtained in group sessions. Since it seemed impractical to obtain reaction-time and confidence measures during group sessions, these measures were obtained during a final block of trials during which *S*s were run individually.

Throughout the group sessions, each *S* was given 10 cards from each of his eight decks and completed this set of 80 cards before he was given the next set. During each trial, *S* (*a*) selected a card, (*b*) marked his prediction of which criterion would occur, and (*c*) observed the criterion. The *S*s were instructed that their primary task was to attempt to predict correctly the criterion on the back of each card. They were further informed that the only cue to the correct criterion was color. Instructions stressed that sequential dependencies had been eliminated by shuffling. Each of the eight blocks of trials contained four sets of 80 cards. These trial-by-trial predictions provided the measures of *maximizing* and *hit rate* relevant to Problem I and the measure of $P(R)$ relevant to Problem II.

After each block in the group sessions, $P(S)$ relevant to Problem II was obtained by using the subjective probability apparatus described above. The *S* was instructed to partition the total length of the rod into four sections in the same proportion as his subjective probabilities were partitioned for the four alternatives. Obtained in connection with each of the eight criterion variables was the subjective conditional probability distribution of the criterion variable, given each of the cues. Also at the completion of each of the eight blocks, each *S* ranked the eight tasks in terms of difficulty, thus yielding the *difficulty ranking* measure relevant to Problem I.

At the completion of eight blocks of trials, *S*s participated individually in a supplementary block of trials during which *reaction time* and *confidence* in prediction were measured. The following procedure was used for each trial. First, *S* was told which criterion variable would be used. Second, the cue side of an appropriate card was presented to *S* in the reaction-time apparatus, starting the timer. Third, *S* verbally predicted which criterion was on the back of the card, stopping the timer with the voice microphone. Fourth, *S* indicated his confidence in the correctness of his prediction by setting a single marker on the subjective probability apparatus. Fifth, *S* was told which criterion was correct. This procedure was repeated for each of the 320 trials. The instructions stressed the importance of accuracy over speed.

TABLE 2

PARTIAL PRESENTATION OF THE ANALYSIS OF VARIANCE OF THE EFFECT OF THREE UNCERTAINTY MEASURES UPON FIVE DIFFICULTY MEASURES

Source	Hits		Maximizing		Confidence		Reaction Time		Difficulty Ranks (Friedman Test)	
	df	*F*	*df*	*F*	*df*	*F*	*df*	*F*	*df*	χ^2
H(Cue	1	4.89*	1	<1.00	1	3.25	1	1.33	1	3.06
H(Criterion)	1	375.44***	1	53.55***	1	43.85***	1	17.78***	1	16.00***
T(Cue; Criterion)	1	79.33***	1	36.12***	1	11.56**	1	7.17*	1	10.56**

* $p < .05$.
** $p < .01$.
*** $p < .001$.

RESULTS AND DISCUSSION

Problem I

The purpose of Problem I was to discover the relation between each of the four different measures of uncertainty and the five different indicators of inferential difficulty. The significance tests summarized in Table 2[3] include, as sources of variance, the three variables constituting the main effects of the factorial design: *H*(Cue), *H*(Criterion), and *T*(Cue; Criterion). The fourth uncertainty variable, *H*(Criterion | Cue), is not itself included as a source of variance because it is an additive combination of the last two of the included sources. In order to employ a procedure that was both conservative and consistent, each effect was tested by using the Effect × *S MS* as the error term. Because rankings possess only ordinal scalar properties, Friedman tests were substituted for an analysis of variance to evaluate the significance of the effects upon Block 8 difficulty rankings. Friedman χ^2s were calculated individually for each uncertainty variable by first collapsing to median ranks across the other two uncertainty variables.

A common pattern emerged from all five difficulty measures. As shown in Table 2, effects of *H*(Cue) never reached the .01 level of significance; with one exception, effects of *H*(Criterion) and *T*(Cue; Criterion) exceeded the .01 level for all five dependent variables. Figure 1 and Table 3 indicate that the significant effects are strong and that inferential difficulty increases as *H*(Criterion)

TABLE 3

MEAN CONFIDENCE IN ACCURACY OF PREDICTION AND REACTION TIME AS A FUNCTION OF UNCERTAINTY CONDITIONS

H(Cue)	*H*(Criterion)	*T*(Cue; Criterion)	*H*(Criterion | Cue)	Reaction Time	Confidence
Low	Low	High	Low	.739	.820
High	Low	High	Low	.723	.811
Low	Low	Low	Intermediate	.773	.652
High	Low	Low	Intermediate	.825	.630
Low	High	High	Intermediate	.821	.604
High	High	High	Intermediate	.883	.571
Low	High	Low	High	.934	.500
High	High	Low	High	.956	.483

---

[3] A complete description of the analyses of variance and Friedman tests performed can be found in Peterson and Ulehla (1963).

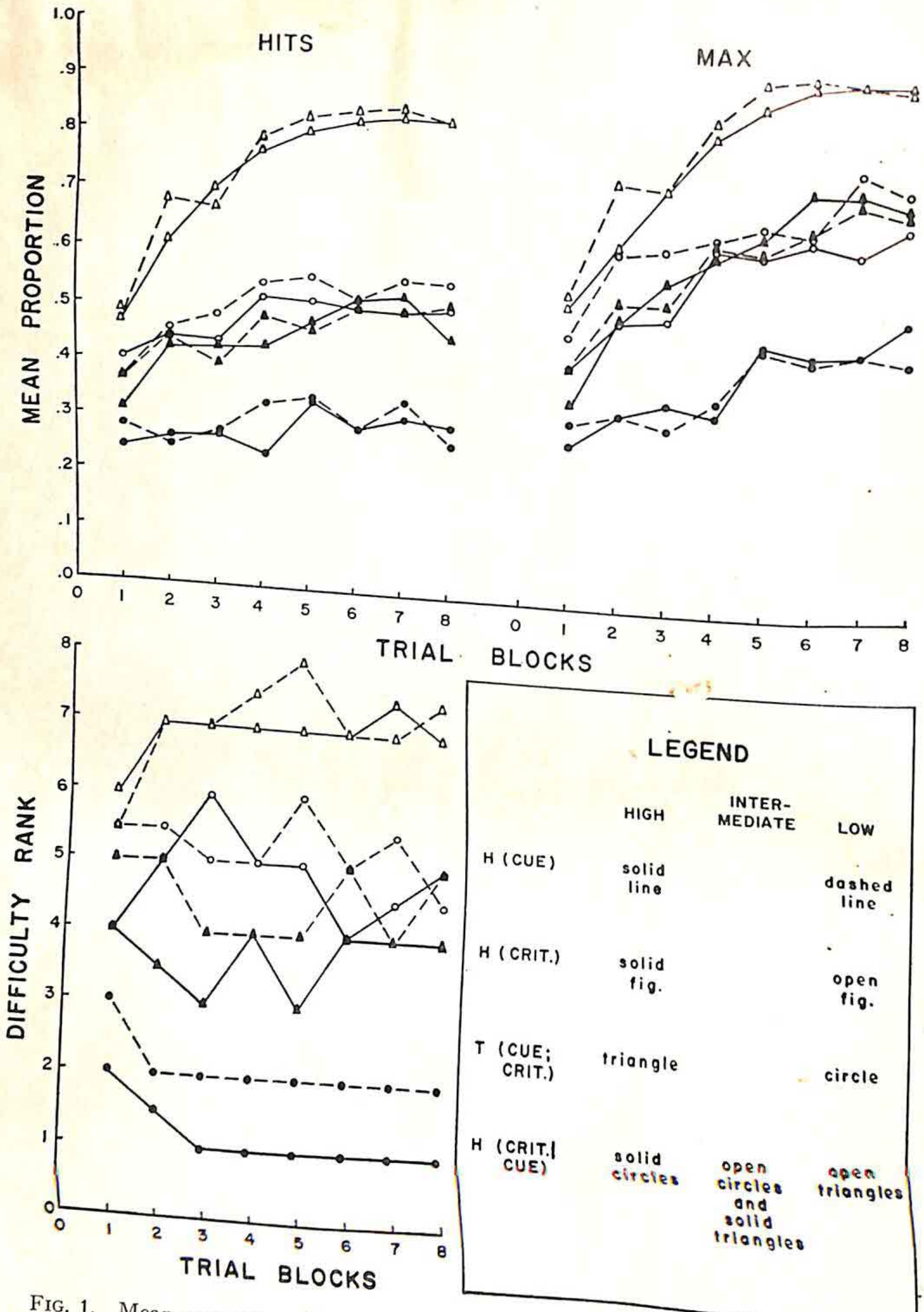


FIG. 1. Mean proportion of hits, mean proportion of maximizing responses, and median ranked difficulty plotted by trial blocks. (Rank 1 is the most difficult and Rank 8 the least difficult.)

increases and/or as T(Cue; Criterion) decreases.

All five difficulty measures, as shown in Fig. 1 and Table 3, segregate themselves into three levels. These three levels correspond to the three levels of H(Criterion | Cue) which resulted from the factorial design. H(Criterion | Cue) is independent of H(Cue), the factor which appeared to have no effect; it is mathematically an additive combination of H(Criterion) and T(Cue; Criterion) and empirically appears to combine the strong effects of these factors.

The lack of effect of H(Cue) is subject to qualification because of its low experimental variation relative to the other uncertainty components. Subject to this qualification, a simple generalization can be drawn from this experiment. Of the four uncertainty measures considered in the introduction, H(Criterion | Cue) is sufficient to account for the variance in inferential difficulty.

The important uncertainty of an inference task does not appear to include the uncertainty concerning which cue will occur, but rather the uncertainty remaining about which criterion will occur after information has been provided by the cue. As this conditional criterion uncertainty increases, S takes longer to make his decision, is less likely to make the optimal prediction, expresses "justifiably" less confidence in his prediction, and concludes that the task is more difficult. The S is also less likely to be correct in his prediction, but this is in part due to the lessened predictability of a more uncertain environment.

Why should the conditional uncertainty of a set of events impair prediction? Several types of explanation are possible. Since increasing the uncertainty of a set of alternatives

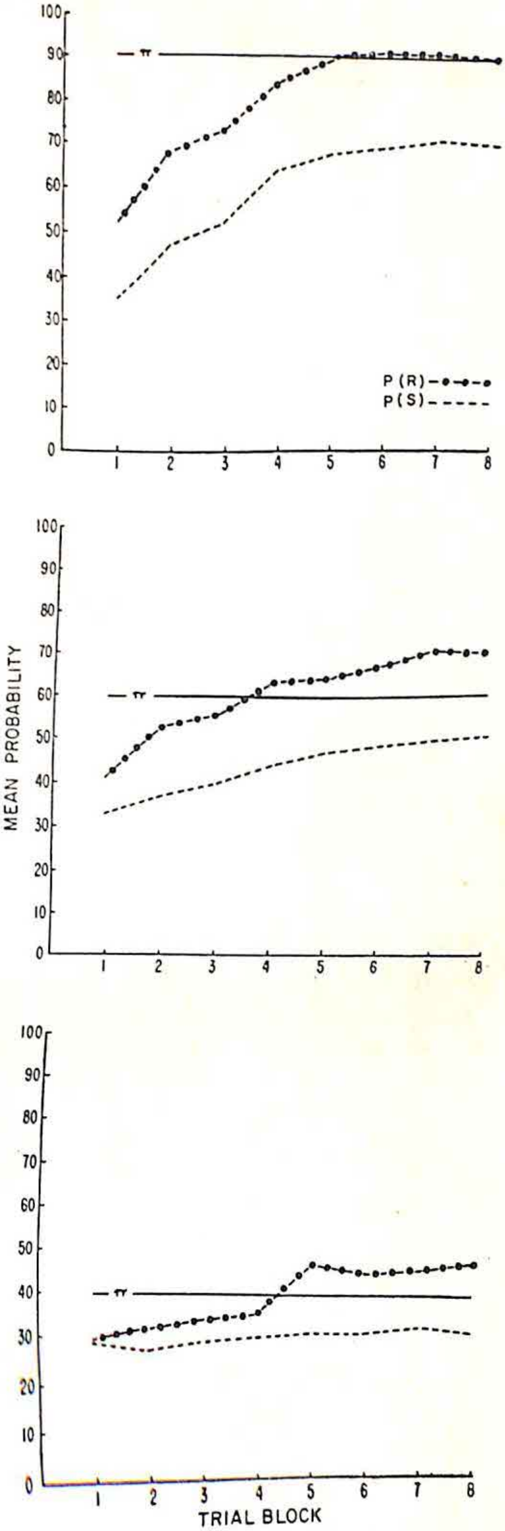


FIG. 2. Change in $P(R)$ and $P(S)$ as a function of trial blocks for π's of .90, .60, and .40.

typically flattens their probability distribution, the discrimination of which alternative is most likely becomes more difficult. Insofar as optimal behavior is learned through reinforcement, learning may be impaired because flattening the probability distribution of the events leads to a smaller differential of reinforcement; that is, less frequent reinforcement of optimal predictions and more frequent reinforcement of suboptimal predictions. Finally, the level of success which is achievable through optimal behavior is limited by π which will typically vary inversely with uncertainty, as in this experiment.

Problem II

The results, as summarized in Fig. 2, confirm the three relations between probabilities suggested in the introduction. First, Block 8 $P(R)$ approximated or slightly exceeded corresponding values of π. Only in the case where $\pi = .60$ did the difference reach even the .05 level of significance ($t = 2.27$). This result extends the frequent probability-learning finding to a more complex experimental situation.

Second, Block 8 $P(S)$ was significantly lower than the corresponding π at each π level (t's, all of which exceed the .01 level of confidence, equaled 3.83, 3.22, and 6.12 for π's of .90, .60, and .40, respectively). This result broadens the empirical base of the generalization that $P(S)$ underestimates π.

Third, $P(S)$ was significantly less than the corresponding $P(R)$ throughout all eight blocks of trials ($F = 65.57$, $p < .001$). This result confirms an implication of the conjunction of the first two findings. Insofar as subjective probability means that probability expressed by S, a methodological implication is that response probability should not be used as an unmodified measure of subjective probability.

Although response probabilities and subjective probabilities are not equal, response probabilities did account for roughly 30% of the variance in subjective probabilities which was not accounted for by event probabilities. A stable average partial correlation of .55 between subjective and response probabilities with event probabilities held constant characterized the last half of the eight blocks of trials. This correlation analysis was based on all conditional probabilities, not only those associated with the most frequent event.

Thus, fairly complex inferential behavior was shown to bear orderly relations to mathematical properties of the environment. This experiment extended two quantitative areas of research, the first area relating information-theory measures of the environment to adequacy of performance, the second area relating conditional probabilities of the environment to those of behavior.

REFERENCES

Attneave, F. Psychological probability as a function of experienced frequency. *J. exp. Psychol.*, 1953, **46**, 81–86.

Attneave, F. Symmetry, information, and memory for patterns. *Amer. J. Psychol.*, 1955, **68**, 209–222.

Brainard, R. W., Irby, T. S., Fitts, P. M., & Alluisi, E. A. Some variables influencing the rate of gain of information. *J. exp. Psychol.*, 1962, **63**, 105–110.

Edwards, W. Probability learning in 1000 trials. *J. exp. Psychol.*, 1961, **62**, 385–394.

Miller, G. A., & Selfridge, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.

Peterson, C., & Ulehla, Z. J. Uncertainty, difficulty, and probability learning. Report No. 31, 1963, University of Colorado, Behavior Research Laboratory.

(Received June 24, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 531–538

GROUPING BASED ON PHENOMENAL PROXIMITY

IRVIN ROCK AND LEONARD BROSGOLE[1]

Yeshiva University

The question was raised whether the Gestalt law of grouping by proximity is based on proximity of stimuli on the retina or perceived proximity in phenomenal space. To tease apart these 2 possibilities an earlier experiment of Corbin was repeated in which the stimulus array is tilted back into the 3rd dimension. By including a measure of constancy for the perceived distance between points it was possible to show in 2 experiments, that the crucial factor is phenomenal proximity. This finding questions the Gestalt assumption that grouping is based on certain objectively given features within the stimulus.

The grouping of stimuli on the basis of proximity is one of the major Gestalt laws of the spontaneous organization of the visual field. Presumably, more proximal stimuli on the retina yield stronger forces of attraction between their corresponding loci of excitation in the brain than less proximal stimuli. However, it has not been demonstrated that it is the anatomical closeness within the proximal stimulus array which governs grouping. The only investigator who ever raised this question was Corbin (1942), who speculated that grouping may be accomplished on the basis of the perceived spatial relations between the points in a stimulus array, and that this phenomenally perceived proximity and retinal proximity, although usually coinciding, can be experimentally separated and thrown into conflict.

Corbin's method was to present an array of illuminated points in the dark seen as columns when in the frontal plane. A gradual lateral tilting of the stimulus configuration into the third dimension progressively foreshortened the horizontal distance between the columns as projected to the retina. Corbin reasoned that under monocular viewing the amount of tilt required to produce a shift from a grouping of columns to one of rows could be predicted in accordance with the actual projection of the image on the retina. Under conditions where the oblique plane of the points was made visible, however, the perceived distance between the columns should remain more or less constant, and a grouping of rows should never be seen despite the severity of stimulus tilt—i.e., if grouping is based upon the *perceived* distance between points.

Since a significantly greater tilt was required to produce a shift in the perceived organization when *O* could see the slant of the array of points, but a shift under these conditions nevertheless did occur, Corbin concluded that both retinal proximity and phenomenally perceived proximity play an equally important role.

The following experiments represent an attempt to expand upon and refine Corbin's procedure. We hypothesized that if grouping is based upon phenomenally perceived proximity, the shift that Corbin obtained might have been due to the falling off of constancy, not to a limitation of the phenomenal factor in grouping. In order to test this hypothesis, we included a measure of constancy at

---

[1] The authors wish to express their appreciation to John Ceraso for his helpful suggestions during the course of this study.

various angles of stimulus tilt, which Corbin had not done.

EXPERIMENT I

Method

Apparatus.—The apparatus consisted of luminously painted glass spheres of $\frac{5}{16}$ in. in diameter, strung vertically and presented in a dark room. They were 3 in. apart vertically and 4 in. apart horizontally so that when in the frontal plane all *O*s would see columns. The spheres were strung to a frame (30 in. in height and 40 in. in width) which could be swiveled about a vertical axis at its right end so that the array could be presented at any angle from frontal-parallel (0°) to 90° if desired. The degree of slant could be read off directly from a protractor mounted to the frame. In order to eliminate the contour formed by the total array of points as a factor which might affect constancy, *S*s viewed the array through a 9-in. diameter circular aperture placed in front of it. Thus, regardless of the angle of the frame behind the aperture, *S* saw the array within a circular region.

Procedure.—Observations were made under monocular conditions and binocular conditions in a totally dark room. The monocular condition was always presented first, since otherwise a memory effect yielding some degree of constancy might have carried over from the binocular condition and spuriously obtruded itself.

The *S* was seated 10 ft. from the array. His head was centered in relation to the circular aperture. Following dark adaptation, and using either his left or right eye, *S* was asked to respond to each stimulus presentation in terms of whether he saw columns or rows. The instructions depicted the experiment as being one of esthetic judgment. It was explained that the luminous dots could be seen to fall either into columns, going up and down, or rows, going across, and that it was *S*'s initial impression to each presentation that was desired. The array was presented at 0, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, and 75° of tilt in a prearranged random order. After each judgment, *S* closed his eyes and the aperture was covered by a shield while *E* changed the angle of tilt. In this way *S* could not see the array being tilted. (Corbin's procedure was to permit *O* to rotate the array until he reported it was just as easy to see columns as rows. This method tends to encourage the operation of sets and expectations —for example, the perceived grouping at the outset may carry over and affect the grouping at a later moment. Also, searching for the point where it "is just as easy to see columns as rows" may entail a difficult judgment about equivalence.) Following the monocular grouping judgments, the array was again presented; now at 0, 25, 35, 45, 55, 65, and 75° of tilt in a prearranged random order and a distance judgment was obtained. The *E* specified two columns (those closest to the axis of rotation) and *O* was to note the horizontal distance between them. The array was then covered and *O* indicated when a $\frac{1}{4}$ in. wide variable luminous horizontal line, located at a distance of 10 ft., appeared equal in length to the column separation. Only one grouping and extent estimate was requested per degree of tilt in each of the two conditions. For the distance judgments, ascending and descending trials were counterbalanced for the various angles of tilt. Following the distance judgments, the entire procedure was repeated using binocular vision.

Subjects.—Nine males and six females ranging in age from 15 to 57 served as *S*s. The majority, however, were high school students. Their mean age was 23.3 yr. All were naive as to the purpose of the experiment.

Results

Despite the random order of presentation, *S*s were consistent in that they did not reverse their reported grouping at angles beyond the point at which they shifted to rows. Under monocular vision *S*s saw the array as columns until, on the average, it was tilted back to an angle of 43.0°. By trigonometric computation, the retinal separation horizontally between points would equal the vertical separation at 41.4°. Hence, the result shows little evidence of any constancy effect, as expected. Under binocular vision, on the other hand, the average position at which the shift occurred was 53.3°, a value significantly different from the monocular value at the .01 level of confidence. Figure 1 shows the number of *S*s seeing a column organization at the different tilts for monocular and binocular vision. It may be noted that whereas almost all *S*s in the monocular condition had switched to rows by 45°, only one *S* in the binocular condition had done so.

The distance measures correspondingly showed a continuous and significant decrease for each successive increment of tilt from 0° to 75° under monocular observation (see Fig. 2). (The average of the *SD*s for the seven different tilts of the array was .47 in.) For binocular observation they did not show any appreciable change until the array was tilted back to 55° or more. (The average of the *SD*s for the seven tilts was .65 in.) (It will be noted that the average perceived distance between columns is 3.1 in.—binocular—or 3.4 in.—monocular—even at 0° whereas the objective distance between column dots was previously stated to be 4 in. The discrepancy is partly due to the fact that the objective distance is given center-to-center whereas the judgments are no doubt made interior-to-interior. The latter distance measures 3.7 in. In addition there is probably some illusion of magnitude involved wherein the thin luminous line appears longer than an equivalent separation between the columns of dots.)

It is clear that with binocular vision *S*s were to some extent able to take slant into account in perceiving the distance between columns. The fact that constancy falls off at 55° suggests that the lateral distance between points is no longer *perceived* to be greater than the vertical distance between them. This would lead to the prediction that a grouping shift would occur at about this point. This is precisely what happened. Hence one cannot say that the shift presents evidence that retinal proximity as a determinant begins to become effec-

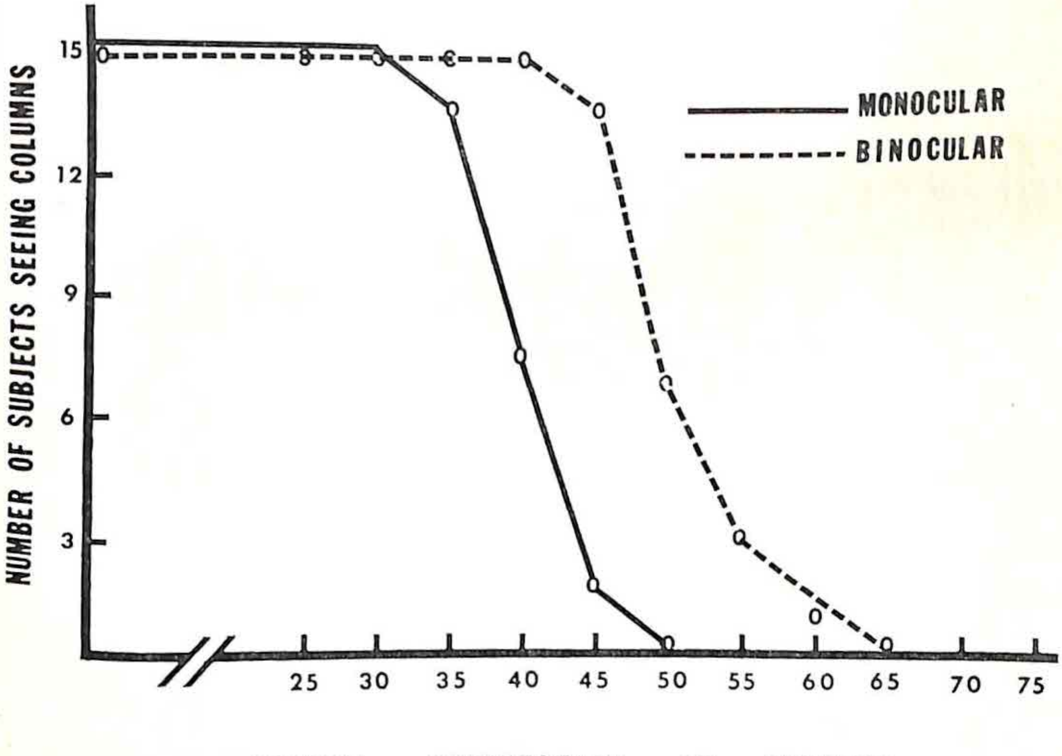


FIG. 1. Number of *S*s seeing columns at each orientation of the array: Exp. I.

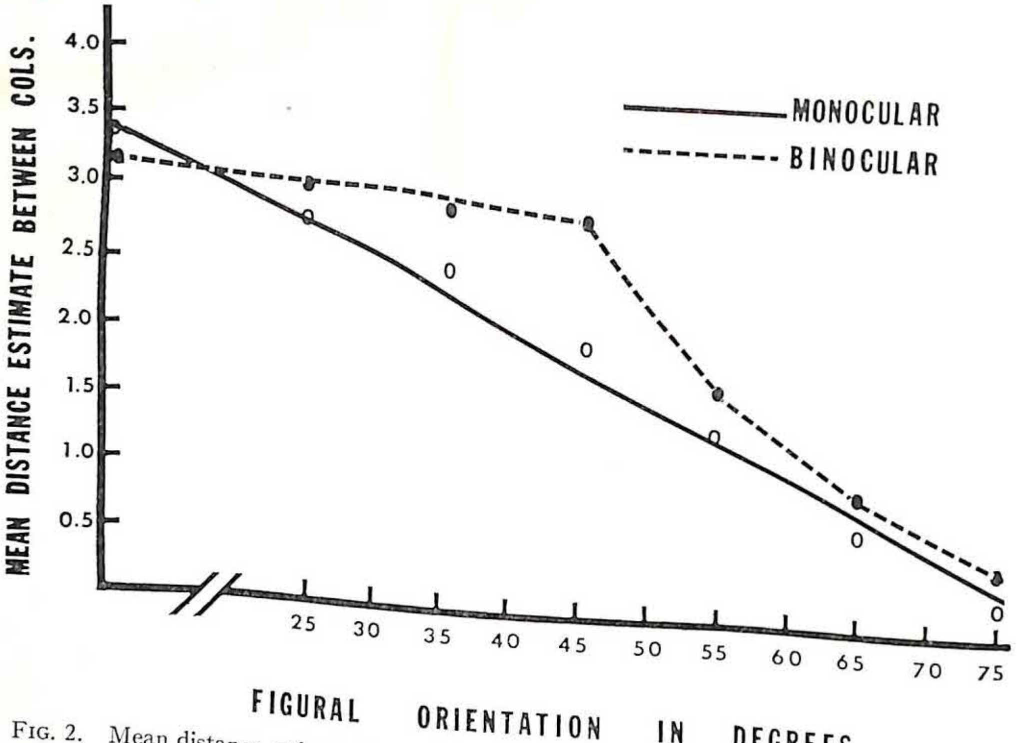


FIG. 2. Mean distance estimate between columns at each orientation of the array: Exp. I.

tive at certain angles of tilt as Corbin contended.

EXPERIMENT II

Method

Procedure.—The procedure was the same as in Exp. I except for the following three modifications:

1. In order to increase the constancy effect still further under binocular viewing, *O*s were permitted to view the array without the circular aperture in front of it.
2. In order to determine at what point the perceived distance between the columns equaled that between the rows, distance judgments were obtained for the vertical separation as well as for the horizontal. This necessitated the adding of an adjustable luminous vertical line to the apparatus. At each angle of tilt, then, a horizontal and vertical distance judgment was obtained. The two judgments for each angle of tilt were separated in time as part of a prearranged randomized order.
3. The order of presentation of the monocular and binocular conditions was counterbalanced between *S*s. Although there was good reason to avoid presenting the monocular condition second in Exp. I, the criticism could be made that the results for the binocular condition were influenced by practice or the like.

Subjects.—Eight naive *S*s, five males and three females, served as *S*s. Their mean age was 25.0 yr.

Results

The effects disclosed by the first experiment were magnified through the adoption of the modified procedure. Whereas under monocular viewing a mean of 41.9° of rotation was required to produce the transformation in figural organization, 62.5° was necessary in the binocular condition. This difference was significant beyond the .01 confidence level. Figure 3 gives the results for grouping.

As to the effect of order, although there only were four *S*s in each subgroup, certain differences seem to be suggested. When the monocular condition was first, the shift occurred on the average at 43.75° while the shift in the binocular condition which followed was at 68.75°. When the binocular condition was first, the shift occurred at 56.25° and the shift for monocular viewing was at 40°. Thus both orders yield a large difference between monocular and binocular viewing, but the monocular-binocular order yields a greater difference. Comparing the two conditions which came first with one another reveals an appreciable difference, but the two conditions which came second show a difference of 28.75° in the average point of shift. It is not clear why the difference is exaggerated when the monocular and binocular conditions are second, but it would seem to be due primarily to the greater effectiveness of binocular viewing when it comes second.

As in the first experiment, constancy was found to prevail only under binocular viewing. Figure 4 gives the results of distance judgments. (The average of the *SD*s for the seven tilts of the array for monocular viewing was .52 in. and for binocular viewing it was .48.) It is clear that under monocular viewing the distances between columns were estimated more or less in accord with visual angle, as was the case in Exp. I. (As a matter of fact the distance judgments under monocular viewing in both experiments fell off at a rate somewhat greater than would be predicted on the basis of visual angle.) Under binocular viewing the distance between columns did not show any appreciable departure from perfect constancy until at least 55°. (In general the tendency toward constancy was stronger in Exp. II than Exp. I for the more extreme tilts.) The Wilcoxon signed-ranks test was used for execut-

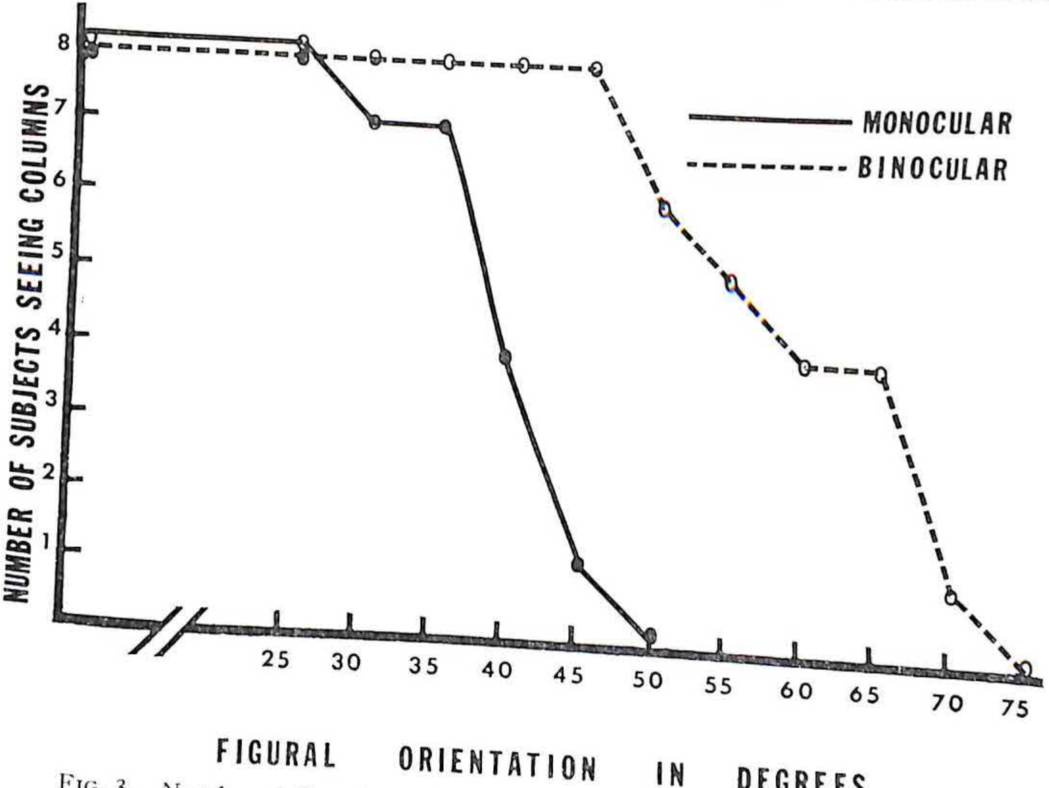


FIG. 3. Number of *S*s seeing columns at each orientation of the array: Exp. II.

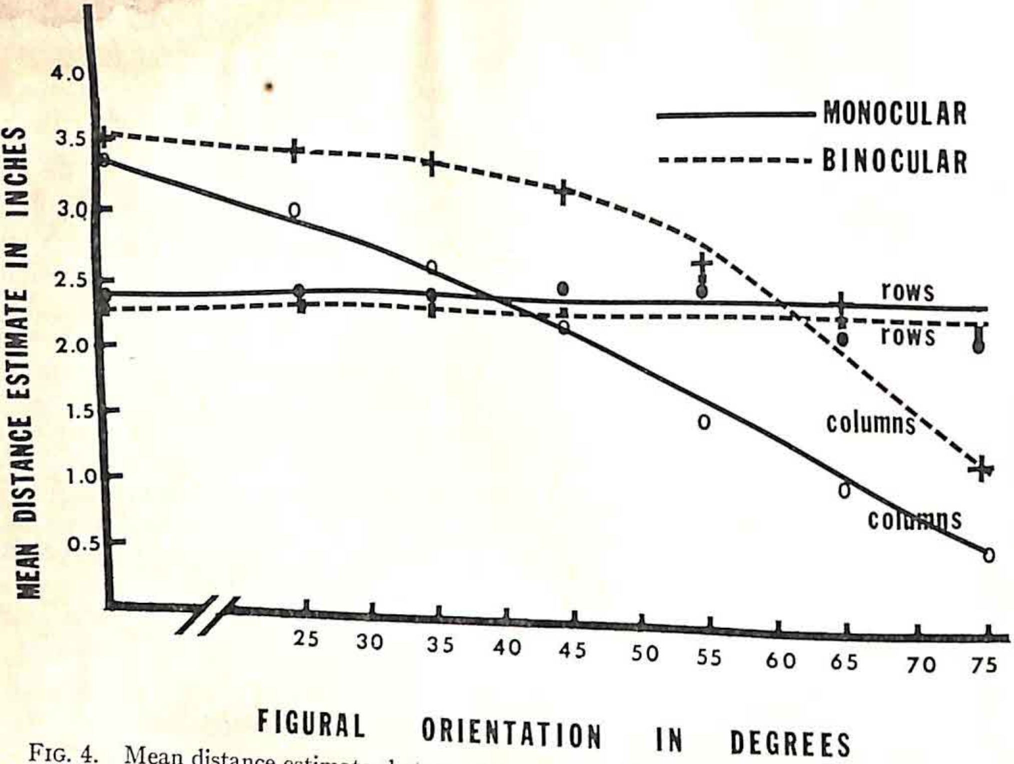


FIG. 4. Mean distance estimates between columns and between rows at each orientation of the array: Exp. II.

ing a series of multiple comparisons following the disclosure of significant main effects by the Friedman two-way analysis of variance. The .01 confidence level was pre-established as the region for the rejection of the null hypothesis. It was determined that under monocular viewing there was a significant decrement in the magnitude of the estimates for all tilts of the array 35° or more from those offered at 0° of tilt. Also there was a significant decrement in the distance estimates for virtually every successive increment in the tilt of the array. On the other hand, only the estimates at 75° of tilt in the binocular condition differed significantly from those at 0°.

The distance between rows for monocular or binocular vision remained fairly constant throughout (note that the average distance is less than the true value of 3 in.) As can be seen, the two monocular curves cross at approximately 40° indicating that at this angle the points were perceived as equidistant in each direction and this corresponds closely with the point of shift in grouping. For binocular vision, the points were, on the average, perceived as equidistant in each direction at 62°, since the two binocular curves cross at about this point. This coincided with the point of shift in grouping. It is of significant import that not one *S* under either of the two conditions had shifted to a grouping of rows so long as the horizontal separation was perceived to be greater than the vertical extent.

To summarize these findings, whereas tilt produced a marked and continued degeneration of horizontal extent judgments in the monocular condition, constancy prevailed in the binocular condition in that the perception of distance was veridical throughout a considerable range of tilts of the array. These differences in the perception of spatial relations coincided with predicted differences in grouping.

DISCUSSION

The results of both experiments are clear in showing that grouping is not based on the retinal distance between points. The principles of grouping play an important role in Gestalt theory. It was on the basis of these autochthonous principles that the field was held to become structured into discrete and segregated entities. Since an attempt was made at offering an alternative to empiricism, it became necessary to specify completely objective unifying factors, such as proximity or similarity. Prior to any experience, then, these objective factors must operate because they determine the way in which the field will be organized. It, therefore, now comes as quite a challenge to this approach to discover that, at least in the case of one grouping factor, it is not the contiguity of elements within the proximal stimulus that governs the perceptual outcome. Rather, grouping appears to be based upon perceived proximity which is itself the outcome of complex organizational processes involving the central integration of information about distance. The fact that Gestalt theorists played such an important role in promulgating the notion of perceptual constancy should not be allowed to confuse the present issue. When the array is tilted, the dots along the rows could be perceived to be farther apart than those along the columns because of constancy, but this does not require that *O* must see columns. In other words, regardless of how the spacing of the dots is perceived, the grouping should still be based on proximity between the elements in the retinal image according to the Gestalt notion of grouping.

We are left with an intriguing problem. Why should grouping occur on the basis of experienced proximity rather than retinal contiguity? One answer is that we learned that stimuli which appear closer together are usually part of a common unit—i.e., grouping by proximity may be based on past experience as suggested by Brunswik and Kamiya (1953). If so, it would be paradoxical in that the grouping principles were intended as an alternative to theories which explain organized perception on the basis of past experience. Of course, the problem remains of explaining how the common unit is itself perceived as a unit in the first place.

Needless to say, it is quite possible that other principles of grouping are also based on phenomenal rather than objective factors. For example, elements which are perceived as similar may be seen as part of one unit regardless of whether the corresponding proximal stimulus elements on the retina are objectively similar. It is not too difficult to imagine how this could be tested. The same is true for certain other grouping principles.

The findings concerning grouping are similar in certain important respects to those reported recently by Rock and Ebenholtz (1962) on stroboscopic movement. These authors raised the question whether the necessary conditions for stroboscopic movement entail change of location of the stimulus on the retina (as had been assumed to be the case) or change in the phenomenal location of the stimulus. They found the latter to be the necessary condition—apparent movement is seen even where there is no change in the locus of retinal stimulation so long as the stimulus is located in two separate places in space. Concerning both studies the question may be raised as to how "phenomenal" proximity or "phenomenal" location can be held to play a causal role in perception. The answer would have to be that the brain processes underlying such experience mediate the perceptual outcome. Nevertheless, it is useful to speak of "phenomenal" factors since there may be various stimulus conditions which would lead to the same outcome provided they share the feature of yielding the same experienced spatial relations. It is these experienced spatial relations—not spatial relations given in the retina—that seem to be crucial. In the case of stroboscopic movement, at least two different stimulus conditions yielded phi, although neither involved change of the retinal locus of the stimulus. Furthermore, if it should turn out to be correct that we have learned to see stroboscopic movement or learned to group by proximity, then it follows that the first step in the process is perceiving certain spatial relations. Hence for the time being, it is convenient to refer to the necessary stimulus conditions in phenomenal terms. Ultimately we shall want to go beyond this way of specifying the conditions.

REFERENCES

Brunswik, E., & Kamiya, J. Ecological cue-validity of "proximity" and of other gestalt factors. *Amer. J. Psychol.*, 1953, **66**, 20–32.

Corbin, H. H. The perception of grouping and apparent movement in visual depth. *Arch. Psychol., N. Y.*, 1942, No. 273.

Rock, I., & Ebenholtz, S. Stroboscopic movement based on change of phenomenal rather than retinal location. *Amer. J. Psychol.*, 1962, **75**, 193–207.

(Received June 24, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 539–544

CONDITIONING OF THE ALPHA RHYTHM IN MAN

R. ALBINO
University of Natal

AND

G. BURNAND[1]
Radcliffe-on-Trent, England

Conditioning of the α rhythm to a tone was attempted in 3 experiments. In 2 of these inspection of the records showed no conditioning. No conditioning was observed in 4 Ss to a kinesthetic stimulus. In a 3rd experiment 8 Ss were given 20 conditioning (CS tone) and 4 extinction trials with control on "attention." Significant suppression of α rhythm occurred, but it was small and variable between individuals. The effect seemed to be due to conditioning to the tone, temporal conditioning, and direct suppression by the tone. It seems that conditioning of the α rhythm can occur but the phenomenon is unstable and highly variable.

The alpha rhythm, appearing in the brain of a man in a dark room, can be suppressed by illumination of the retina. This suppression may, it is believed, be conditioned to a neutral auditory stimulus (Jasper & Cruikshank, 1937; Loomis, Harvey, & Hobart, 1936; Travis & Egan, 1938). The rhythm may, however, be suppressed by attention or by an instruction to solve some simple problem in mental arithmetic. It also fluctuates spontaneously in many individuals.

This fluctuation and sensitivity to attentive mental states compels caution before assuming a stimulus to have become a conditioned suppressor of alpha rhythm. Any apparent conditioning might be no more than suppression due to anticipation of the stimulus, surprise at its appearance, or at the sudden withdrawal of the illumination on the test trials.

Shagass and Johnson (1943) attempted to control anticipation by telling *S* whether the light would appear or not. Whatever the information given—even if false—the apparent conditioning persisted. It seemed as if anticipation was unimportant for conditioning.

Individuals vary greatly in their susceptibility to conditioning of the alpha rhythm. Shagass and Johnson (1943) in one experiment conditioned 6 of 12 *S*s, and in another (Shagass, 1942) obtained the response in 7 of 8 *S*s.

Even if the response is established it seems to be unstable. Loomis et al. (1936) only observed the effect a few times; Travis and Egan (1938) found only 34% of stimulus presentations to cause suppression; and Jasper and Shagass (1941a, 1941b) also found the response to be unstable. Not only is it unstable but it appears weak—Shagass (1942) found it necessary to use an index (.01 × Alpha Index × Amplitude Index) which would magnify any suppression present, a procedure inclining one to believe the effect to be small. Knott (1941), after careful examination of the evidence and experiments of his own, concluded that there was only tentative evidence for conditioning. Since these early studies there have been others. These are well summarized in the 1955 Marseilles colloquium. All suffer from the defect that only selected trials are exhibited, and there is little evidence to show that the observed phenomena are stable and reproducible. For example, both Gastaut et al. (1957) and Mundy-

---

[1] The authors wish to thank R. Russell for providing facilities for this work at University College, London.

Castle (1958) failed to confirm results of Rousinov and Smirnov (1957) which, from the reports, appear obvious and reproducible.

It is clear that the phenomenon has not been demonstrated beyond doubt. Nor has the role of other factors (such as attention) known to suppress the rhythm been determined. The following experiments were designed to help resolve some of this uncertainty. They attempt to control attention, use a simple measure of alpha rhythm (mean amplitude over a given period), and provide further information on the strength of conditioning, on its variation between individuals, and on its modification by attention.

Three experiments were performed. One is reported in full, the others are merely commented upon. Some experiments with individuals under various conditions are also discussed.

METHOD

Subjects.—All *S*s were university students and most had some training in psychology. Those selected had a consistent α rhythm of high amplitude, easily and surely suppressed by light.

Apparatus.—Alpha rhythm was recorded on a four-channel Ediswan EEG. One channel was used to record alpha rhythm from left occipito-parietal electrodes. A second channel recorded potentials from electrodes on the left zygoma and bridge of the nose. These potentials were taken as an index of eye movements. Other channels recorded stimulus presentations. In all experiments *S* reclined semisupine in a chair.

Procedure

Experiment I.—A moderately quiet 500 cycles/sec tone was presented for 5 sec. to *S*. During the last 3 sec. of this period, an illuminated 6-v. bulb was exhibited just above eye level and 6 ft. away. Two *S*s were used and each was given about 40 trials.

Experiment II.—Four *S*s in each of the three groups were given 40 conditioning trials, the intervals between trials varying randomly from 5 to 25 sec. The *S*s of Group I were instructed to relax, and those of Group II to attend to the sound stimuli. For Groups I and II the stimuli were ordered as in Exp. I.

Group III was given a rubber bulb and asked to squeeze it on command for 10 sec. This formed the CS. The UCS began 5 sec. after *S* was told to squeeze the bulb and ended when he released it. (Both CS and UCS were recorded parallel to the EEG.) The same *S*s composed Groups I and III.

To estimate sensitization to the tone in Groups I and III, 10 trials with tone alone were given before the trials began and again after the twentieth conditioning trial.

Experiment III.—The CS was a tone (500 cps, through headphones) lasting 6 sec., and the UCS a light, just above eye level, illuminated for 6 sec. beginning 1 sec. after the onset of the tone. Each *S* had 20 conditioning trials and 4 extinction trials (tone alone) separated by an intertrial interval of 23.5 sec. under each condition, far and near.

Both stimuli were controlled by a mechanical timing device. The stimuli, parieto-occipital EEG, and eye-muscle potentials were recorded.

There were two groups of four *S*s each. Group I was instructed to look at the light and Group II "that they need not look at the light." Each *S* was tested under two conditions: far (f), with the lamp 6 ft. away from his eyes and near (n), with the lamp 9 in. away. To control order effects the n and f conditions were alternated (Table 1).

The aim of this design is to control attention and factors related to it. The instruction to Group I to look at the light presumably increases their attentiveness. Looking at a near light will, perhaps, on account of the need for sudden voluntary convergence, further increase attention. Thus, all Group I *S*s may, if attention assists conditioning, be expected to condition more easily, and this effect should be more marked under the near condition. Conditioning should occur less easily in Group II, and least of all under the far condition. It should be remarked, however, that attention may not be the only factor controlled. Both the increased luminance of the near light and proprioceptive feedback from the muscles of convergence may also affect the α rhythm.

RESULTS

Inspection of the records in Exp. I (one *O*) and in Exp. II (two *O*s) showed no visible evidence of conditioning.

Because of the absence of visible evidence of conditioning in the records of Exp. I and II, the records of Exp. III were measured in order to disclose even small degrees of suppression. Each record of 20 conditioning and 4 extinction periods was divided into consecutive 2-sec. periods, and the mean alpha amplitude determined for each of these. In a sequence of four such 2-sec. periods, the first would contain the first second of the CS and the second before this, and the three subsequent periods would cover the UCS (light) and the remainder of the tone. The mean amplitude in the intertrial periods was determined in the same way.

Evidence of conditioning.—If conditioning occurred, then the ratio of the mean α amplitude on the first extinction trial to that observed in the preceding and succeeding intertrial periods should be less than unity. Table 1 exhibits these ratios.

In 13 out of 16 instances the ratio is less than unity and it may be assumed that conditioned suppression of alpha rhythm did occur in response to the tone alone ($p = .02$, two-tailed binomial test). However, except in a few instances the suppression is small in amount.

An analysis of variance of the table shows no significant differences between either Groups I and II or Cond. f and n, though the difference between Groups I and II approaches significance at the 5% level. It may be the case, therefore, that Group II conditions more easily, which is contrary to the expectation that poor attention will diminish the likelihood of conditioning. Also, most conditioning was expected in Group I with the light near, whereas, in fact, less occurred under this than under the other conditions. It seems that attention inhibits—or masks—conditioning.

The observed suppression of α rhythm on the test trials could be attributed to a sensitization or similar effect of the tone alone. Such an effect would perhaps show itself by brief suppression near to the onset of the CS, whereas the suppression due to conditioning would be expected to appear later and nearer the point in time at which the UCS normally is presented. Thus, one would expect, if conditioning were present, to find a low α amplitude in that part of the extinction trial when the light would normally have been on as compared with intertrial α amplitude. To test this supposition the mean α amplitude over the first 2 sec. of the "light-on" period and that during the preceding 2 sec. intertrial and the first 4 sec. of the following four intertrial periods were measured.

This was done for the first and last two extinction trials. These measures were compared with a hypothetical ranking which would be expected if conditioning had occurred—one in which the intertrial periods were assigned equal high ranks and the test

TABLE 1

RATIOS OF MEAN ALPHA RHYTHM ON FIRST EXTINCTION TRIALS TO THAT ON INTERTRIAL PERIODS

Group	S[a] and Order	Light		Difference (f − n)
		Far	Near	
I	1. f, n	0.920	1.193	−.273
	2. f, n	0.357	0.225	.132
	3. n, f	0.852	1.025	−.273
	4. n, f	0.656	1.160	−.504
II	5. f, n	0.305	0.342	−.037
	6. f, n	0.846	0.764	.082
	7. n, f	0.713	0.517	.196
	8. n, f	0.654	0.391	.263

[a] This column is identical in Table 2 and is, therefore, omitted from Table 2.

TABLE 2

SUPPRESSION OCCURRING LATE IN THE CS

Group	Light		Sum
	Far	Near	
I	−15	−12	−27
	18	22	40
	16	4	20
	18	−24	−6
II	20	40	60
	−1	2	1
	7	42	49
	30	10	40

periods equal low ranks. The *S*[2] values from this procedure are given in Table 2 and have a mean value of 36 (3 test periods, 12 intertrial periods). The result for Group II is significant (p = .05, two-tailed t test), showing that suppression occurs in this group in the period where the light would have normally been on. In Group I the α suppression, if present, is occurring mainly near the onset of the CS.

It remains to consider the occurrence of blocking in this period near to the onset of the tone. It may be due either to temporal conditioning effects or to some sensitization factor or anticipation resulting from the change in stimulus conditions on the extinction trials. ("Will the light come on or not?")

It may be assumed that if temporal conditioning has occurred, then any suppression will tend to appear before and soon after the onset of the tone, whereas anticipation effects would lead to suppression near to the beginning of the period when the light would normally have been on. Thus, suppression of the mean α amplitude in the first second of the tone taken together with the preceding intertrial second may indicate the extent of temporal conditioning (and the effect of the tone itself), whereas suppression during the .5 sec. of the tone immediately prior to the light may indicate anticipation effects.

To test these possibilities the average α amplitude during the last .5 sec. of the tone prior to the light-on period, and that during the first second of the tone and the second preceding it was measured over the last four conditioning trials and the extinction trials.

These measures were compared with a hypothetical ranking in which the conditioning and first extinction trials were assigned a rank of 3, the subsequent extinction trials being ranked 6, 7, and 8, respectively. This ranking assumes extinction, and therefore gradual reappearance of α rhythm to occur over the last three extinction trials.

The *S* values obtained were barely significant for both the 2-sec. and the later .5-sec. period. However, all were positive, and the overall sum for the 2-sec. period exceeds that for the .5-sec. period. There is, therefore, slight evidence for temporal conditioning.

A further test for temporal conditioning in the period preceding the CS was made. The mean α amplitudes of the nine 2-sec. intertrial periods preceding the last conditioning trial and the first extinction trial were ranked. These ranks were compared with a hypothetical ranking, such that negative *S* values indicated a decrease of α rhythm with time. The overall *S*s were negative but barely significant. However, the values correlated −.63 with the mean α amplitude in the .5 sec. of the tone immediately preceding the "light-on" period, and −.35 with that for the first second of tone and the immediately preceding second. Thus it seems that suppression early in the tone is partly due to temporal conditioning, as well as to the tone itself and to anticipation. This evidence is slight, but is sufficient to demand caution before assuming the observed suppression during the tone to be entirely due to direct conditioning to the sound.

---

[2] The *S* values refer to the values obtained by using Kendall's procedure for computing the rank correlation coefficient (Siegel, 1956).

It remains to consider how far the instruction did, in fact, encourage attention or inattention. The *S*'s own reports do not settle this question, but they make it clear that all in Group I voluntarily converged on the light in both far and near conditions, and that they waited expectantly for it to appear in order to do so. Those in Group II did not converge voluntarily, but there is, of course, no way of excluding the possibility that they attentively anticipated the appearance of the stimuli.

Individual Experiments

1. With the stimuli timed as in Exp. I, and the light 9 in. from the eyes, a series of 40 conditioning trials was given. The mean alpha rhythm for the second .5 sec. of the tone was measured. In the first 20 conditioning trials the α rhythm at the onset of tone exceeded that before onset, whereas in the last 20 trials it tended to be less. The change was significant at the 5% level.

2. Using a loud tone with conditions as in Paragraph 1 above no consistent changes were observed in 28 trials in the α rhythm before and after the onset of the tone; there was, that is, no tendency for the loud tone itself to suppress the rhythm.

3. An attempt was made to establish temporal conditioning by exhibiting the light for 2 sec. at randomly distributed intervals of either 8, 10, or 12 sec. Forty-eight trials were given, 24 with the light far (6 ft. away) and another 24 with it near (9 in. away).

Temporal conditioning was observed by inspection just after the light had been brought near. The experiment was repeated on two other *Ss* with a constant time interval between stimuli of 10 sec. In both inspection showed obvious temporal-conditioning effects.

DISCUSSION

These experiments tend to support the belief that suppression of the α rhythm may be conditioned. However, the effect is small and shows considerable individual variation, even though under appropriate conditions it was observed in all eight *Ss* in Exp. III. In these eight *Ss* the depressed α rhythm varied from .4 to .9 of its original resting value. Only in two instances was it depressed by more than 50%.

The effect of attention seems to be to decrease the ease of conditioning, and, if this is associated with convergence of the eyes, conditioning is entirely abolished. (With Subject 2 this was not so. This individual conditioned exceptionally well while apparently converging and attending.)

There is some slight evidence that the observed effect is a mixture of direct conditioning to the tone, temporal conditioning, and suppression. It is not possible on the basis of these experiments to decide on the relative importance of these factors.

In general, these experiments support the view of Knott (1941) that there is only tentative evidence for conditioning of the α rhythm to tone.

REFERENCES

GASTAUT, H., JUS, C., MORRELL, F., VON LEEUWEN, W. S., DANGIER, S., NAQUET, R., REGIS, H., ROGER, A., BEKKERING, D., KANEP, A., & WERRE, J. Etude topographique des réactions électroencéphalographiques conditionées chez l'homme. *EEG clin. Neurophysiol., Amsterdam*, 1957, 2, 1–34.

Jasper, H., & Cruikshank, R. M. Electroencephalography: II. Visual stimulation and the after image as affecting the occipital alpha-rhythm. *J. gen. Psychol.*, 1937, **17,** 29–48.

Jasper, H., & Shagass, C. Conditioning occipital alpha rhythm in man. *J. exp. Psychol.*, 1941, **28,** 373–388. (a)

Jasper, H., & Shagass, C. Conscious time judgments related to conditioned time intervals and voluntary control of the alpha rhythm. *J. exp. Psychol.*, 1941, **28,** 503–508. (b)

Knott, J. R. Electroencephalography and physiological psychology: Evaluation and statement of problem. *Psychol. Bull.*, 1941, **38,** 944–975.

Loomis, A. L., Harvey, E. N., & Hobart, G. Electrical potentials of the human brain. *J. exp. Psychol.*, 1936, **19,** 249–279.

Mundy-Castle, A. C. An appraisal of electroencephalography in relation to psychology. *J. Nat. Inst. Personnel Res. monogr. Suppl., Johannesburg*, 1958, No. 2.

Rousinov, U. S., & Smirnov, G. A. Electroencéphalographiques de l'activité nerveuse supérieur. In M. H. Fishgold & H. Gastaut (Eds.), Conditionnement et réactivité en électroencéphalographique. *EEG clin. Neurophysiol. Suppl., Amsterdam*, 1957, No. 6.

Shagass, C. Conditioning the human occipital alpha rhythm to a voluntary stimulus: A quantitative study. *J. exp. Psychol.*, 1942, **31,** 367–379.

Shagass, C., & Johnson, E. P. The course of acquisition of a conditioned response of the occipital alpha rhythm. *J. exp. Psychol.*, 1943, **33,** 201–209.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Travis, L. E., & Egan, J. P. Conditioning of the electrical response of the cortex. *J. exp. Psychol.*, 1938, **22,** 524–531.

(Received June 25, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 545–552

EXTINCTION OF THE HUMAN EYELID CR AS A FUNCTION OF THE DISCRIMINABILITY OF THE CHANGE FROM ACQUISITION TO EXTINCTION[1]

KENNETH W. SPENCE, M. J. HOMZIE, AND E. F. RUTLEDGE

University of Iowa

2 experiments investigated rate of extinction of the eyelid CR in 105 humans as a function of the degree of discriminability of the procedural changes that occur with shift from acquisition to extinction. Experiment I found that following continuous reinforcement at a CS-UCS interval of 500 msec. Ss extinguished more quickly when the UCS was omitted than when it was delayed 1,500 or 2,500 msec. Experiment II revealed that Ss run in the context of a masking situation (probability learning) were much more resistant to extinction than under standard conditions. Rate of extinction was negatively correlated with degree of recognition of stimulus changes with introduction of extinction procedures.

Previous studies from the Iowa laboratory (McAllister, 1953; Reynolds, 1958) have shown that performance during extinction following partial reinforcement in conditioning is, as predicted from the Hull-Spence drive theory, higher under the condition in which the drive level (*D*) is maintained by extending the CS-UCS interval (2,500 msec.) than that in which it is not (UCS discontinued). Contrary to this prediction, however, no difference is obtained under these extinction procedures following continuous reinforcement (Reynolds, 1958; Spence, Rutledge, & Talbott, 1963). In the latter instances, moreover, the extinction is extremely rapid, the group curves decreasing to a random or operant level of response after two to three nonreinforced trials, with many individuals never giving a response after a single nonreinforcement.

This evidence of extremely rapid extinction following continuous reinforcement, plus the fact, obtained by questioning, that under these conditions, all *S*s readily recognize the change in conditions with shift from acquisition to extinction, led us to theorize that with the shift most *S*s recognize that they have been conditioned and, thereupon, immediately adopt an inhibitory set not to blink to the CS. In accord with this interpretation is the well-established finding that extinction is not nearly as rapid following partial reinforcement as after continuous. Under a partial-reinforcement schedule *S* has already experienced failures of reinforcement on some portion of the trials (e.g., 50%) and, hence, the shift to 0% reinforcement is not as readily discernible as is the shift from 100% reinforcement to 0%. Presumably, such an inhibitory set as we have postulated would not develop in extinction following partial reinforcement, at least not immediately. Hence, the performance of partially reinforced *S*s in the early stages of extinction would not be affected by such an inhibitory set and, thus, would not be expected to exhibit the precipitous decrement that occurs following continuous reinforcement.

---

[1] This study was carried out as part of a project concerned with classical conditioning in humans under Contract Nonr-1509(04) between the University of Iowa and the Office of Naval Research.

This hypothesis as to the operation of an inhibitory set that depends upon the recognizability of the changes that occur with shift to extinction also nicely accounts for the finding that the difference in performance level under the two extinction procedures, omission of the UCS and lengthening of the CS-UCS interval, is contingent upon the original conditioning having been conducted under partial reinforcement. In extinction following partial reinforcement, the absence of the inhibitory set during the early phase of extinction would mean that the CR would not be affected by this interfering factor and, hence, its effective strength ($\bar{E}$) during this period would be a function of the developing effects (I_n) of nonreinforcement and the level of drive (D) present. Since the drive level of the group in which the UCS is omitted would be less than the group in which the UCS is present but delayed, its performance should be lower. In extinction following continuous reinforcement, on the other hand, *S*s under both extinction procedures would immediately recognize the shift to extinction and adopt the inhibitory set. Immediate disappearance of the CR would thus be expected under both extinction procedures with the result that there should be no difference in their performances.

The two experiments described in this report attempted to investigate further the implications of this theorizing as to the relation of performance during extinction to the discriminability of the experimental conditions holding in the acquisition and extinction periods. Both attempted to vary the discriminability in the direction of reducing it so that the likelihood that *S*s would discern the shift to extinction and adopt the inhibitory set would be minimized. In the absence of the inhibitory set extinction should be retarded, the response decrement depending instead on the gradual development of I_n with successive nonreinforcements.

EXPERIMENT I

The first variation in procedure was introduced on the basis of some exploratory observations of Goldstein (1962) which indicated that the shift to extinction with the extended CS-UCS interval was not as readily recognized when the CS duration was the same during extinction as in conditioning. Whereas, in previous studies the CS duration was increased from 550 msec. in acquisition to 2,550 msec. during extinction, in the present experiment CS duration was 2,550 msec. in both periods, thus reducing their discriminability. During extinction Group I had the UCS discontinued, Group II had the CS-UCS interval extended from 500 msec. to 2,500 msec., while Group III had this interval increased to 1,500 msec. The shorter interval used with Group III was introduced as a further attempt to reduce the discriminability of acquisition and extinction conditions.[2]

Method

Subjects.—A total of 113 students from an introductory course in psychology was employed. Of these, 8 were eliminated, 4 because they failed to reach an acquisition criterion of at least five CRs in the last 20 trials of acquisition, 3 for apparatus failure or *E*'s error, and 1 who gave CRs to test-trial presentations of the CS alone. Of the 105 remaining *S*s, 30 were classified as voluntary-form responders (V responders) on the basis of the form of their anticipatory eyelid response (Spence & Ross, 1959). The somewhat larger than usual proportion of such V responders resulted, in part, from the fact that the criterion was changed slightly to include not only *Ss* who gave 50% or more such voluntary-form responses over all acquisition trials, but also *Ss* who gave such responses on 50% or more of their last 20 acquisition trials, i.e., Trials 31–50. These latter *Ss*, 9 in number, developed this form of response late in the conditioning. Since they were responding in the same fashion as the remaining V responders at the end of acquisition it was thought best to include them in this group. The extinction data of these V responders will be presented as well as the data for the remaining 75 *Ss* who did not give many voluntary-form responses. The latter *Ss* will be referred to as C responders.

---

[2] Two unpublished investigations (Myers, 1950; Rutledge, 1962) from this laboratory have found that the increase in percentage of CRs with a CS-UCS interval of 1,500 msec. is very small, especially in the interval 150–500 msec. following the onset of the UCS.

Apparatus.—The apparatus was essentially the same as that used in previous studies from this laboratory (cf. Spence, 1953). The CS, a 70-db., 500-cps tone generated by a Hewlitt-Packard oscillator, had a duration of 2,550 msec. The UCS was a 1-psi air puff of 50 msec. duration delivered to the right eye through an orifice .062 in. in diameter. A CR was recorded whenever the Brush recording pen showed a deflection of 1 mm. or more in the range 150–500 msec. following the onset of the CS. Responses in the interval 0–150 msec., which were relatively infrequent, were classified as original responses and were not included in the data.

Conditioning procedure.—Described as an experiment concerned with study of attention and fatigue, *S* was instructed to fixate upon and keep looking at a 6-cm. circular milk-glass disc located directly in front of him. He was also instructed to be as relaxed as possible and not to attempt to control the reactions of his eye. Following the reading of the instructions each *S* was given three presentations of the CS alone followed by a single presentation of the UCS alone. Immediately following these preliminary trials, 49 conditioning trials were given in which the CS-UCS interval was 500 msec. A photoelectric film programmer predetermined intertrial intervals of 9, 12, and 15 sec., averaging 12-sec. intervals for each *S*. No ready signal was employed.

Extinction procedures and experimental design.—As described above, *Ss* were extinguished under three different nonreinforcement procedures. On the fiftieth trial and for 30 subsequent trials, the UCS was omitted for *Ss* in Groups I-C and I-V, the CS-UCS interval was lengthened to 2,500 msec. for Groups II-C and II-V and to 1,500 for Groups III-C and III-V. Twenty-five *Ss* were assigned to each of the groups containing C responders and there were 10 *Ss* each in Groups I-V, II-V, and III-V.

Results

C responders.—Extinction curves for the C responders are presented in the left-hand graph of Fig. 1. The curves give the percentage of anticipatory responses occurring in the interval 150–500 msec. following the CS as a function of the number of previous nonreinforced trials. The values at the point 0 on the abscissa of the graph give the performance levels on the last 10 acquisition trials (41–50). The last 10 *Ss* in each group of C responders were assigned in such a manner as to equate roughly their final acquisition levels.

As may be seen from the graph, the C responders who had the UCS discontinued (Group I-C) extinguished very rapidly. After only two nonreinforcements the percentage of anticipatory responses was at a random level. On the other hand, extinction was considerably slower in the case of *Ss* who had the UCS delayed either 2,500 or 1,500 msec. (Groups II-C and III-C). The response curves for these groups drop more slowly and do not reach the 10% level by 30 trials. The performance (per cent CRs) of Group I-C was significantly below that of Groups II-C and III-C over the first 10 extinction trials, the Mann-Whitney *U* test (two-tailed) giving a p value of .003 for both comparisons. Over all 30 trials the difference between Groups I-C and II-C was not significant ($p = .17$), whereas that between I-C and III-C was ($p = .003$). The extinction performances of Groups II-C and III-C did not differ significantly either for the first 10 trials or over all 30 trials.

V responders.—Extinction curves for the V responders are presented in the right-hand graph of Fig. 1. The extinction function and terminal acquisition level (Trials 41–50) are presented in the same manner as for the C responders. As may be seen, the

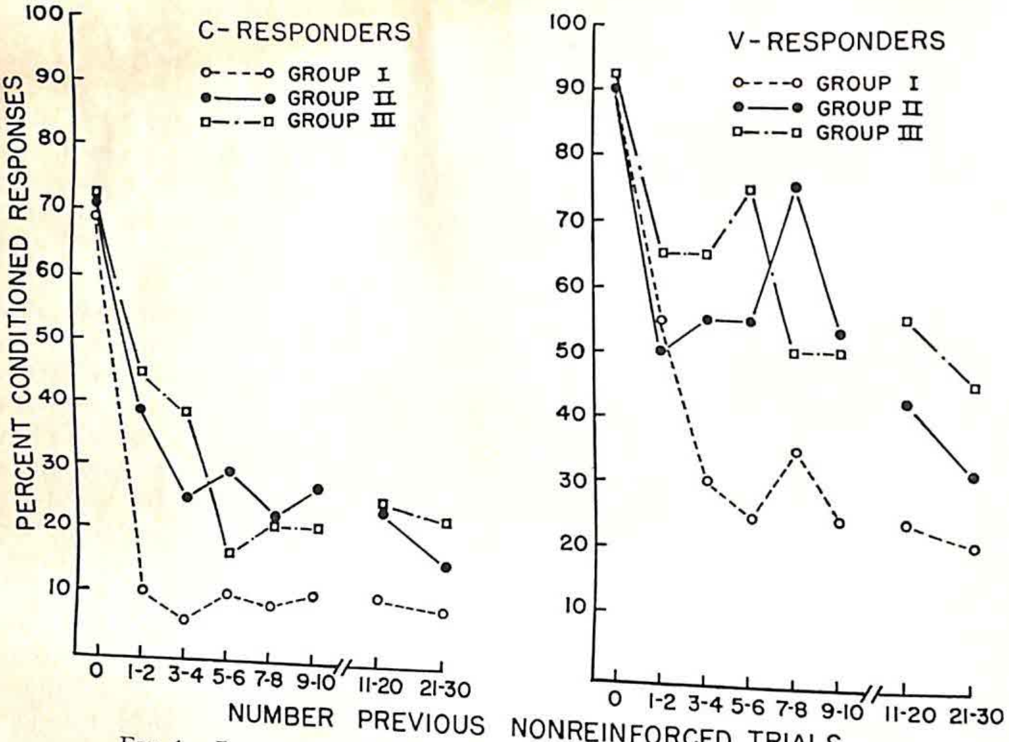


FIG. 1. Percentage of CRs during extinction as a function of number of previous nonreinforced trials.

V responders who had the UCS discontinued (Group I-V) extinguished more rapidly than did *S*s who had the UCS delayed either 2,500 or 1,500 msec. The performance of Group I-V was significantly below that of Groups II-V and III-V both over the first 10 trials and over all 30 trials, the Mann-Whitney *U* test (two-tailed) giving a p value of .009 for both comparisons made over the first 10 trials, and p values of .032 and .021 for Group I-V vs. Group II-V and Group I-V vs. Group III-V, respectively. The extinction performance of Groups II-V and III-V did not differ significantly either over the first 10 or all 30 trials.

C responders vs. V responders.—Comparison of the final acquisition levels at Point 0 in Fig. 1 reveals that the V responders attained a higher final level of conditioning performance than did the C responders. The mean percentage of responses during the last 10 acquisition trials was 91.0 for the V responders and 70.6 for the C responders. The difference between these means was highly significant, $t\ (103) = 3.99,\ p < .001$.

Comparison of the extinction performances of the C and V responders requires that the differences in their final acquisition levels be taken into consideration. Following a method recently suggested by Anderson (1963) we first obtained for each *S* the following index of extinction performance.

$$\text{Index} = (\text{Rext.} - \Sigma R_n/N)/(\text{Rext.} - \text{Racq.})$$

Where Rext. is the asymptotic extinction level (here assumed to be 10% for all *S*s in both groups), Racq. is the final acquisition level for each *S* (per cent CRs in last 10 acquisition trials), and ΣR_n is the number of CRs given over N extinction trials. This index gives a measure of performance during extinction relative to the difference between the final acquisition level and the assumed asymptotic extinction level. The higher the index value, the greater the resistance to extinction or the slower the extinction. Comparison of these index values for *S*s in Groups II-V and III-V combined with those for *S*s in the combined groups II-C and III-C revealed the former to be higher, the differences between their distributions being significant at the .002 level for Extinction Trials 1–10 and at the .001 level for Trials 1–30 (Mann-Whitney *U* test). That is, the V responders extinguished significantly slower than the C responders when the UCS was delayed during extinction. In the case of *S*s extinguished without the UCS, however, the difference in extinction rate as determined by comparing the distribution of these indexes for Groups I-C and I-V was not significant ($p < .06$ for Trials 1–10 and $p < .60$ for Trials 1–30).

EXPERIMENT II

This experiment involved primarily a comparison of the rate of extinction performance of Group II-C of the first experiment with a group of *S*s (IV) that was conditioned and extinguished under conditions that were designed to minimize knowledge on the part of *S* that he was in a conditioning experiment. Questioning of *S*s in a recent study of extinction (Spence et al., 1963) after they had completed the experiment revealed that 52% of the *S*s were able to recognize that it was concerned with conditioning of the eyeblink response. This percentage is to be compared with findings of earlier studies from our laboratory which showed that when conditioning but not extinction was involved, only from 5 to 20% of the *S*s were able to identify the experiment as concerned with conditioning. Obviously, the added experience of extinction greatly increased the recognition of the true purpose of the experiment, thus increasing the possibility that cognitive factors such as the inhibitory set we have postulated might play an important role in determining performance during extinction.

We were seeking, then, a technique that would preclude *S* from recognizing that he was participating in a conditioning experiment even when extinction was attempted. It was also hoped that the procedure would decrease still further the likelihood that the changes in the stimulus events relative to extinction could be observed. The technique hit upon was that of imbedding the conditioning experiment within the setting of another learning situation, the probability-learning task as used by Estes and Straughan (1954).

Method

Subjects.—Twenty-nine *S*s from an introductory course in psychology were run concurrently with those in Exp. I. Of these, four were eliminated, two because they met the criterion of being V responders and two because they failed to reach the acquisition criterion of at least five CRs in the last 20 acquisition trials.

Apparatus.—The apparatus was the same as that described in Exp. I, except for the equipment involved in the probability-learning task. This equipment consisted of three signal lights, two push buttons located one at the end of each chair arm, and two Hunter event counters that were activated by the push buttons. One signal light, a 6-cm. circular milk-glass disc that increased in brightness from .004 to 2.95 mL. was located directly in front of *S*. The other two lights, located 8 in. to either side of this central light, were .5-in. (diameter) circular discs covered with translucent paper. The brightnesses of these lamps approximately matched that of the central light.

Conditioning and extinction procedure.—The only difference in procedure between Group II-C of Exp. I and Group IV was the addition of the probability-learning task. The *S*s were given a set of instructions to the effect that the experiment was concerned with the effects of distraction on performance in a difficult problem-solving situation. They were instructed that their task was to predict, on the onset of the center signal light, which of the two small lamps was going to light up and to signify their prediction by pressing the push button located on the left or right arm of their chair. They were told, further, that distracting stimuli in the form of a tone and air puff to their eye would be given in between their response of pressing a button and the lighting up of one of the small lamps. The onset of the 2,550-msec. CS (the tone) coincided with the turning off of the signal light. On reinforced trials the onset of the UCS (puff) followed the onset of the CS by 500 msec. On nonreinforced trials the UCS was presented during the last 50 msec. of the CS duration. Five hundred milliseconds after the termination of the CS one or the other of the small lights came on and remained on for 1 sec. As in Exp. I, 50 conditioning and 30 extinction trials were given.

Results

Figure 2 presents the extinction data for Group IV and Group II-C from Exp. I. As may be seen from the values for Point 0 on the abscissa, Group IV *S*s attained a somewhat lower final acquisition level than did *S*s conditioned under the usual conditions. The difference of 16.4% was significant, providing a t value of 2.56 ($p < .02$). While significant in the present instance, we do not yet know whether this finding is typical. The results of two subsequent studies suggest that conditioning performance under the masking situation is somewhat lower, but not as much as obtained here.

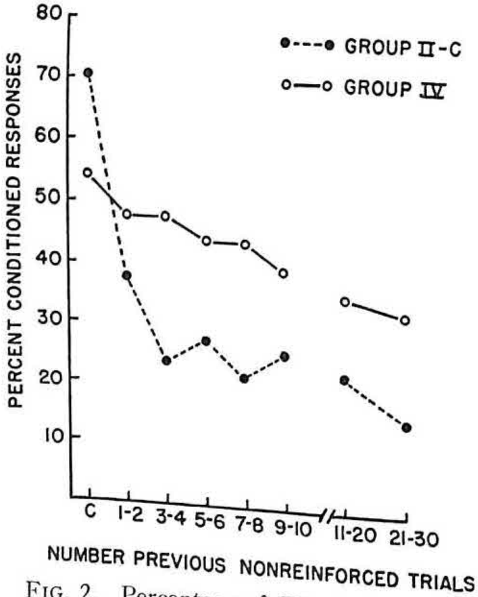


FIG. 2. Percentage of CRs during extinction as a function of number of previous nonreinforced trials.

TABLE 1

PERCENTAGE OF *S*s IN EACH GROUP THAT RECOGNIZED CHANGE IN UCS

Group	C Responders	V Responders
I	84	90
II	76	80
III	48	40
IV	4	—

Especially striking, however, is the relatively slow extinction of Group IV. As may be seen, the precipitous decrement in response frequency of CRs evident in the first 10 trials of all groups in Exp. I is completely absent. Thus, by the ninth and tenth trials, the curve for this group decreased only 14% from the final acquisition level; whereas the corresponding decrease for Group II-C (45%) was more than three times as large.

The number of extinction trials was too small to permit an estimate of the final asymptotic level of responding under the conditions of the masking situation from the group curve. However, further examination of the extinction data showed that the lowest third of the *S*s in terms of resistance to extinction reached the 10% level of responding in Trials 7–10, while the middle third was down to 11% in their last two trials (29–30). These findings suggest that if extinction had been carried out a sufficient number of trials, the group curve would eventually have dropped to the 10% level.

If we assume that the asymptotic level of response of Group IV is the same as for Group II-C, i.e., 10%, the index suggested by Anderson (1963) may be used to compare the rate of extinction to this level of the two groups. Computation of the index gave a higher distribution of values (slower extinction) for Group IV than Group II-C. The difference between the distributions was significant by the Mann-Whitney *U* test at the .005 level for Trials 1–10 and at the .001 level in the case of Trials 1–30.

DISCUSSION

In the case of *S*s in which the UCS was maintained during extinction, Exp. I attempted to reduce the recognizability of the shift from acquisition to extinction by arranging for the CS duration to be the *same* (2,550 msec.) in both phases rather than increasing from 550 msec. to 2,550 msec. as in the previous studies of Reynolds (1958) and Spence, Rutledge, and Talbott (1963). While this change in procedure did lead, in contrast to the findings of the previous studies, to a significant difference between the extinction performances of *S*s extinguished with and without the UCS following continuous reinforcement, there is not complete agreement between these extinction data and the percentage of *S*s in each group that recognized the change in the extinction conditions.

As may be seen from Table 1, the percentage of C responders who, at the end of the experiment, reported recognizing the change in the UCS, either its omission or its delay, was higher for *S*s who had the UCS omitted (Group I) than for *S*s who had the UCS maintained (Groups II and III). This finding is in accord with the suggestion that there is a relation between this recognition and extinction performance. However, the magnitude of the differences among the groups in percentage of *S*s recognizing the change, which were significant ($p < .05$) as determined by chi-square test between I and III and between II and III, but not between I and II, are not in complete accord with the differences in extinction performance among the three groups.[3] If such recognition plays an important role in extinction, the difference between Groups I and II should have been slight, while Groups II and III should have been very different. This was not the case, either for the C or V responders, as the extinction curves of Fig. 1 attest.

A possible explanation of this discrepancy may be that the recognition data do not indicate the point (trial) during extinction that the discrimination takes place. If Group II *S*s recognized the change much later on the average than Group I *S*s, then they would be expected to exhibit the greater resistance to extinction that they did. According to this assumption, questioning of *S*s after a small number of extinction trials, e.g., five, should lead to a significant difference in the proportion of *S*s in these two groups (I and II) that would report the discrimination.

A final bit of evidence on this relation of discrimination to extinction behavior is provided by the data in Table 1 for Group IV. The fact that only one of the *S*s (4%) run in the context of the masking situation reported observing the change in interval between the tone and puff adds further strong support to the negative relationship between such recognition and persistence of the CR in the extinction period.

The extremely slow rate of extinction exhibited by these *S*s is in striking contrast to previous findings with humans. It is, interestingly enough, more comparable to the relatively slow extinction of the eyelid CR found in dogs (Marquis & Hilgard, 1936), monkeys (Marquis & Hilgard, 1937), and rabbits (Schneiderman, Fuentes, & Gormezano, 1962), in whom such higher-order inhibitory sets presumably would not be present.

Taken together the findings of these two experiments appear to be in agreement with the notion that discrimination of the extinction conditions from those holding during acquisition plays an important role in determining the subsequent rate of extinction of the CR. It should be noted, however, that this particular version of the discrimination hypothesis differs from previous formulations of this type in that it introduces as a critical new factor the concept of *inhibitory set*. According to the present hypothesis, discrimination or discernment of the change is assumed to lead to the adoption on the part of *S* of this set not to blink. Extinction is extremely rapid under such circumstances because the inhibitory set results in interference with the CR. However, with conditions arranged so as to preclude or delay the discrimination, the decrement in the frequency of the CR should be more gradual, being dependent upon the cumulative development of inhibition of nonreinforcement (I_n). It should be noted that the present form of this discrimination hypothesis is able, not only to account for the different rates of extinction following continuous reinforcement obtained in Exp. I under the conditions of omitting and delaying the UCS, but also the previously well-established fact of greater resistance to extinction following partial than after continuous reinforcement.

---

[3] As may be seen from Table 1, comparable differences were found among the V responders. However, due to the small number of *S*s, none of the differences were significant.

In concluding this discussion, it is pertinent to call attention to the point that the present version of the discrimination hypothesis with its assumption of an inhibitory set is not conceived as applying to animals either in aversive or appetitionally motivated behavior situations. The hypothesis is put forward primarily to deal with the extremely rapid extinction (one or two nonreinforced trials) that occurs only in humans on the initial extinction of a CR. Furthermore, as the senior author has indicated elsewhere (Spence, 1960, p. 110) the nature of the mechanisms lying behind inhibition in the case of learning situations involving appetitional needs and reward incentives is probably very different from that operating in the extinction of responses learned under aversive motivation. Certainly the empirical facts differ in a great many respects.

REFERENCES

Anderson, N. H. Comparison of different populations: Resistance to extinction and transfer. *Psychol. Rev.*, 1963, **70**, 162–179.

Estes, W. K., & Straughan, J. W. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Goldstein, H. The effects of set factors and reinforcement schedules on the acquisition and extinction of the conditioned eyelid response. Unpublished doctoral dissertation, State University of Iowa, 1962.

McAllister, W. R. The effect of eyelid conditioning of shifting the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 423–428.

Marquis, D. G., & Hilgard, E. R. Conditioned lid responses to light in dogs after removal of the visual cortex. *J. comp. Psychol.*, 1936, **22**, 157–178.

Marquis, D. G., & Hilgard, E. R. Conditioned responses to light in monkeys after removal of the occipital lobes. *Brain*, 1937, **60**, 1–12.

Myers, J. An experimental investigation of the effects of varying the time between the onsets of the conditioned and unconditioned stimuli on the conditioned eyelid response. Unpublished doctoral dissertation, State University of Iowa, 1950.

Reynolds, W. F. Acquisition and extinction of the conditioned eyelid response following partial and continuous reinforcement. *J. exp. Psychol.*, 1958, **55**, 335–341.

Rutledge, E. F. The effect of a dynamic CS on the acquisition of the conditioned eyelid response at an extended CS-UCS interval. Unpublished master's thesis, State University of Iowa, 1962.

Schneiderman, N., Fuentes, I., & Gormezano, I. Acquisition and extinction of the classically conditioned eyelid response in the albino rabbit. *Science*, 1962, **136**, 650–652.

Spence, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

Spence, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

Spence, K. W., & Ross, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

Spence, K. W., Rutledge, E., & Talbott, J. Effect of number of acquisition trials and the presence or absence of the UCS on extinction of the eyelid CR. *J. exp. Psychol.*, 1963, **66**, 286–291.

(Received June 28, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 553–559

STIMULUS OVERLAP IN A MASSED-TRIAL SITUATION[1]

ROSE GINSBERG

Queen's University

7 groups of human Ss were taken through a sequence of massed trials under a 2-choice, noncontingent reinforcement procedure. Experimental overlap between successive trials was introduced, the amount of overlap varying between groups from 1% to 100%. Stimulus overlap due to massed trials was taken into account by modifying the basic learning model of statistical learning theory to allow for the assumption of nonindependent trials. The modified model described the experimental data with considerable success and throws light upon some of the difficulties which have arisen in the applications of statistical learning theory. In particular it was shown that when trials are massed and stimulus overlap is taken into account, the same learning parameter cannot be estimated both from the difference between conditional probabilities of specific responses given the immediately preceding reinforcement, and from the mean learning curve.

It is a working hypothesis in statistical learning theory that stimulation of *S* is independent from trial to trial. Experimental evidence, on the other hand (Straughan, 1956), indicates that stimulus overlap is a factor, at least in massed-trial learning. In addition, Estes (1959) describes certain inconsistencies in the application of statistical learning theory and suggests that these may be a function of the assumption of trial independence under massed-trial conditions. In particular he refers to the two-response, massed-trial experiment where estimates of the learning parameter, θ, from two different sources commonly fail to agree. Both these estimates—one obtained from the equation for the mean learning curve, the other from the difference between conditional probabilities of a given response on trials following the two types of reinforcing events—should, theoretically, yield the same value.

In the present experiment nonindependence of successive trials is deliberately introduced, a predetermined proportion of the visual signal initiating each trial persisting from one trial to the next. A modification of the basic learning model of statistical learning theory to allow for assumptions of nonindependence of trials, is presented, and the data of the experiment interpreted from the point of view of the modified model. In particular the implications of the model in respect to the disparate learning parameter estimations described above, are examined.

Theory.—The stimulus-sampling model developed by Estes and Burke has been described elsewhere (Burke & Estes, 1957; Estes & Burke, 1953). In essence it is assumed that learning involves a sampling process. The experimental situation defines a population of stimulus elements from which, on each trial, *S* draws a random sample. It is this latter, independent sampling assumption, which is modified in the present model by the introduction of two conditional-sampling probabilities θ_s and θ_p, which are the sampling probabilities of an element on Trial n given that it was, or was not, respectively, in the sample on Trial $n - 1$. The introduction of two sampling probabilities follows Straughan's (1956) suggestion that stimulus elements sampled on a given trial are more likely to be resampled on the following trial. And the specific theoretical hypothesis, basic to the present experiment, is that under massed-trial conditions $\theta_s > \theta_p$.

---

[1] This study represents part of a dissertation submitted to the Department of Psychology of Indiana University in partial fulfillment of the requirements for the PhD degree in psychology. The writer is greatly indebted to C. J. Burke under whose direction the investigation was conducted.

In a two-choice noncontingent situation where two responses A_1 and A_2 are available and the probability of reinforcement of A_1 is π, p_n (the probability of an A_1 response on the nth trial) can be expressed as the proportion of elements in the sample conditioned to A_1. In the case of the modified model this proportion has two sources, θ_p comes from the conditioned elements left in the population on the preceding trial (N_{n-1}) and θ_s comes directly from the preceding sample (s_{n-1}); i.e.,

$$p_n = \frac{\theta_p N_{n-1} + \pi \theta_s s_{n-1}}{s_n}.$$

If s_n and N_n are expressed recursively and the solutions of the resultant difference equations substituted in this statement, it can be shown that:

$$p_n = \frac{(1 - \theta_p)^{n-1}(1 - \theta_s + \theta_p)}{1 - (\theta_s - \theta_p)^n} \times (p_1 - \pi) + \pi \quad [1]$$

From the empirical evidence both θ_s and θ_p represent small values so that $(\theta_s - \theta_p)^n$ approaches zero very rapidly as n increases, in comparison with $(1 - \theta_p)^{n-1}$. When $(\theta_s - \theta_p)^n$ is neglected, the equation, like the "independent" sampling model, describes a negatively accelerated curve running from the initial probability to the final asymptotic value of π.

In the present study the signal initiating each trial for S was made up of a number of small parts, and these took on the role of the stimulus elements of the model. Briefly the signal was made up of a potential population of 100 parts and the trial-by-trial presentation of these "elements" exactly followed the assumptions of the model. On Trial n the signal was composed of a predetermined proportion of signal elements from Trial $n - 1$, plus a proportion of elements not in the signal on the previous trial. To preserve the analogy to the model the proportion of the signal persisting from one trial to the next is designated θ_s^* and the proportion that does not originate in the preceding trial signal, θ_p^*.

Apart from the trial-to-trial change in the signal, the experiment was two-choice, with S required on each trial to predict which of two exhaustive and mutually exclusive events, E_1 and E_2, would occur. The two events were not contingent upon S's responses but were presented with fixed probabilities π and $1 - \pi$.

METHOD

Subjects.—There were 252 *S*s, 36 in each of seven groups, drawn from introductory psychology courses at Indiana University.

Apparatus.—The darkened experimental room contained four S booths facing, and equidistant from, a black screen. Each booth contained a response lever and two reinforcing lights. The E sat behind a simple overhead projector, inserting on each trial a punched IBM standard eight-word card. Light traveling through the rectangular punches in the card and reflected onto the screen, provided the signal to which *S*s responded.

Both the projector lamp and the reinforcing lights were automatically controlled, and the time of presentation of signal and reinforcement light, together with S's responses, were automatically recorded on each trial. A trial lasted 5 sec.

The signal.—The sequence of punched cards which provided the signal on each of the 240 trials was made up in the following way. Of the potential punch positions on the IBM card, 100, arranged in a 10 × 10 square,

TABLE 1

PROBABILITIES USED TO CONSTRUCT SEQUENCE OF SIGNALS FOR EACH GROUP TOGETHER WITH SAMPLE SIZE BEFORE AND AFTER STABILIZATION

Groups	θ_p	θ_s	Trials										Function
			1	2	3	4	5	6	7	8	9	10	
I	0.1	0.1	10[a]										Control for II
II	0.04	0.63	4	6	8	9	10[a]						Experimental
III	0.1	0.9	10	18	24	30	34	38	40	42	44	46[a]	Experimental
IV	0.76	0.1	76	26	59	37	52	41	49	44	46[a]		Experimental
V	0.46	0.46	10	18	24	30	34	38	40	42	44	46[a]	Control for III
VI	0.46	0.46	76	26	59	37	52	41	49	44	46[a]		Control for IV
VIIa	Constant		10[a]										General Control
VIIb	Signal		46[a]										General Control

[a] Indicates stable sample size and first trial of occurrence.

comprised the population of signal elements at the outset of the experiment. On Trial 1, $\theta_p{}^*$ of the 100 positions was chosen at random and an IBM card, punched at the selected positions, provided the signal. On each of the remaining 239 trials, $\theta_s{}^*$ of the positions punched on Trial n was randomly chosen for inclusion in the card on Trial $n + 1$, plus $\theta_p{}^*$ of those positions not punched on Trial n.[2] In this way a constant, predetermined, proportion of the signal persisted from one trial to the next.

Each of six groups was assigned a different pair of values for $\theta_p{}^*$ and $\theta_s{}^*$. Three random signal series were constructed for each group to balance out any specific series effect.

Design.—On each of the 240 trials, *S* was required to predict which of two reinforcing lights would go on. A signal initiated each trial and *S* indicated his choice by moving a lever up or down in the direction of one or other of the lights. Presentation of the correct light terminated the trial. For all *S*s the probability of reinforcement of an A_1 response (the position, up or down, counter-balanced across *S*s) was .75. Three random reinforcement sequences were used such that each signal series was run under every reinforcement sequence. For Group VII the signal was constant throughout the 240 trials.

Table 1 lists the various groups and indicates their functions and special characteristics. It will be noted that for every experimental group ($\theta_s \neq \theta_p$) there was a control group matched in respect to the size of the signal on each trial but with $\theta_s = \theta_p$.

To amplify the hypothesized positive correlation between samples on adjacent trials θ_s was substantially greater than θ_p in two of the experimental groups (Groups II and III). Negatively correlated samples were introduced in Group IV where $\theta_p > \theta_s$. Group VII, with constant signal, was intended to provide a basis of comparison with the experimental groups and a check of the appropriateness of the experimental situation in comparison with previous experiments. Because the signal size for all experimental groups stabilized at either 10 or 46 elements, half of Group VII was run with constant signal of 10, half with 46, rectangles of light. The IBM cards used as the slide for these subgroups were chosen randomly from the sequences of signals of appropriate size in the variable-signal groups. A different card was used for each group of four *S*s but within each group the same card was used over all trials.

Procedure.—Each *S* was seated in a booth and instructed to guess on each trial which of the two reinforcing lights in his booth would go on. He was told that at the beginning of each trial a slide projected on the black screen would be the signal for him to make his guess and that he should then move the response lever in his booth up or down towards the chosen light, after which the correct light would go on.

Four *S*s were run at one time and the total sequence of 240 trials took exactly 20 min.

RESULTS

As there was no significant difference between the mean number of A_1

---

[2] Preparation of the sequence of stimuli and the punching of the IBM cards were carried out by means of a program written for the IBM 650 at Indiana University by Richard Millward of Brown University.

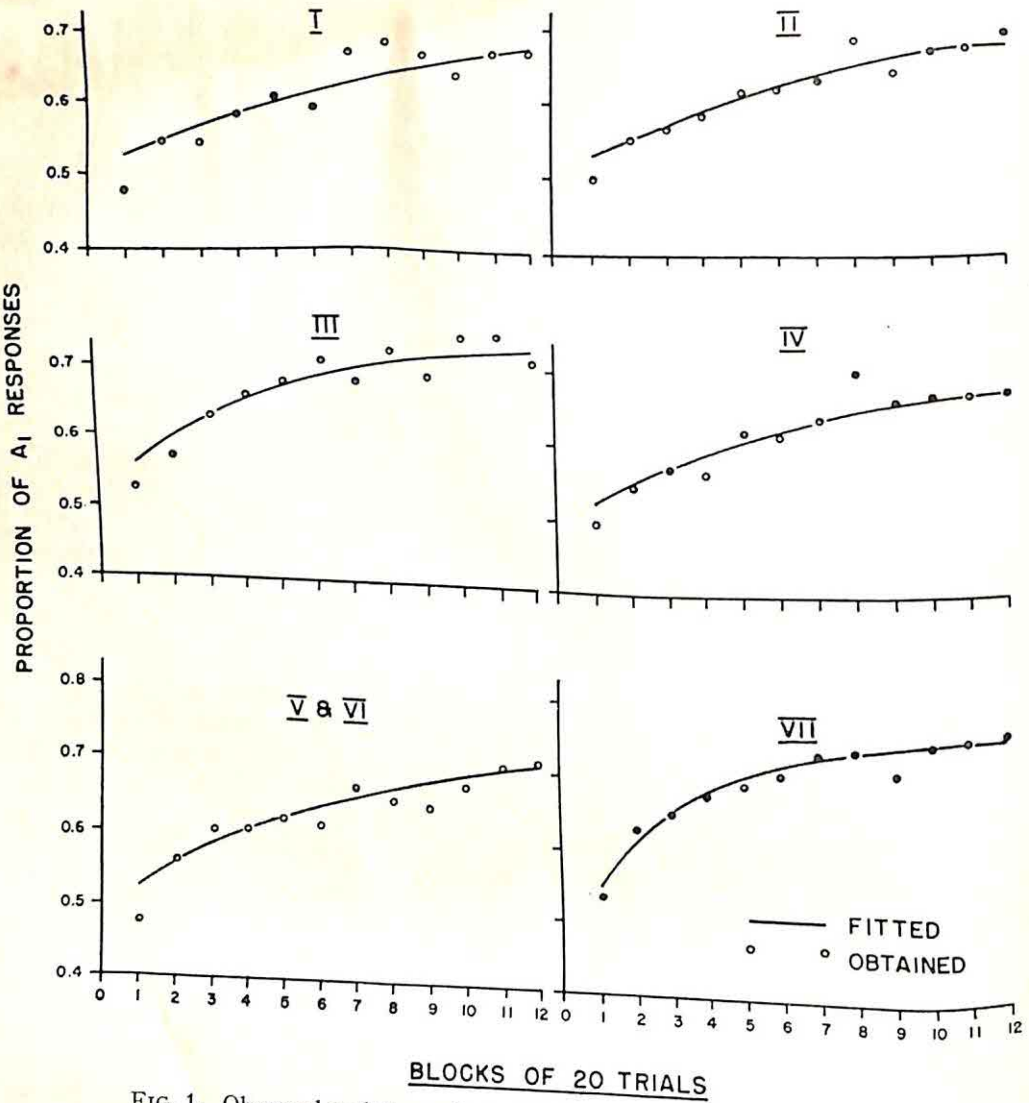


FIG. 1. Observed and theoretical (fitted) proportions of A_1 responses per block of 20 trials for Groups I–VII.

responses over all trials for Subgroups VIIa and VIIIb, t (34) = .16, these data were combined. Groups V and VI differed only in the number of rectangles of light in the signal over the first 10 trials. t tests of the difference of mean number of A_1 responses over the first 20 trials, t (70) = .11, and between the total number of A_1 responses, t (70) = 1.16, were both nonsignificant and again the data were combined.

The data for each group in proportions of A_1 responses per block of 20 trials, are presented in Fig. 1. Inspection of the learning curves in that figure suggests that the groups with variable signal have not yet stabilized at asymptote—certainly none achieve, in 240 trials, the predicted asymptote of .75. For Group VII (constant signal) on the other hand, the learning curve presents a familiar negatively accelerated curve reaching a stable asymptotic level of .75 over the last 60 trials.

Two analyses of variance were performed with respect to (*a*) total A_1 responses over all trials and (*b*) A_1 responses during the last 20 trials. Over all trials only the F for groups was significant ($p < .01$). In the last 20 trials there were no significant differences between groups. These results, together with simple inspection of the curves of Fig. 1, suggest quite strongly that the learning curves are converging in the later trials. And in fact, various other statistical results which are described elsewhere (Ginsberg, 1959) support this conclusion.

Parameter evaluation.—The "overlap parameter" θ_s, can be evaluated from the difference between two conditional probabilities—the probability of an A_1 response given the reinforcing event, E_1 or E_2, on the previous trial, i.e., given an E_1 on Trial $n - 1$:

$$p_n(A_1|E_1) = \frac{\theta_p N_{n-1} + \theta_s s_{n-1}}{s_n} \quad [2]$$

Given an E_2 on Trial $n - 1$:

$$p_n(A_1|E_2) = \frac{\theta_p N_{n-1}}{s_n} \quad [3]$$

Taking the difference between Equations 2 and 3 it can be shown that:

$$\lim_{n} \frac{1}{n} \sum_{v=1}^{n} [p_v(A_1|E_1) - p_v(A_1|E_2)] \rightarrow \theta_s \quad [4]$$

As major changes in s_n are confined to the early trials, θ_s was estimated by summing the difference between the average conditional probabilities over Trials 11–20 and subsequent blocks of 20 trials. The mean of the sum was the estimate of θ_s.

To evaluate θ_p: The expected number of A_1 responses over the fixed n trials, $E(R_n)$, is equal to the sum of the probabilities of an A_1, over the n trials. $E(R_n)$ is empirically determined and the probability of an A_1 response is given by Equation 1. Assuming that $(\theta_s - \theta_p)^n$ can be neglected after, at most, 10 trials, $E(R_n)$ can be calculated over the last 230 trials and the resultant equation solved for θ_p. The precise equation is:

$$\frac{230\pi - E(R_{230})}{\pi - p_1} = \frac{(1 - \theta_s + \theta_p)}{\theta_p} [(1 - \theta_p)^{10} - (1-\theta_p)^{240}] \quad [5]$$

The theoretical curves presented in Fig. 1 have been calculated with parameters estimated by the methods described above. The theoretical curves provide quite good fits for the data, for all groups.

Proportion of elements sampled.—The single parameter θ of the Estes-Burke "independent" sampling model, represents the expected proportion of elements in the sample. The analogous value in the present model is $\lim_{n} s_n/N$. Letting $\lim_{n} s_n/N = \alpha$ it can be shown that:

$$\alpha = \frac{\theta_p}{1 - \theta_s + \theta_p} \quad [6]$$

Table 2 presents the value of θ of the "independent" sampling model for each group, together with θ_p,θ_s, and α estimated by the methods described.

TABLE 2

ESTIMATED PARAMETERS OF TWO MODELS FOR EACH GROUP

Group	Modified Model			"Independent" Model
	θ_p	θ_s	α	θ
I	.0049	.0626	.0052	.0076
II	.0054	.0906	.0059	.0080
III	.0094	.1593	.0110	.0123
IV	.0052	.0581	.0054	.0077
V & VI	.0065	.0470	.0069	.0084
VII	.0190	.0790	.0202	.0203

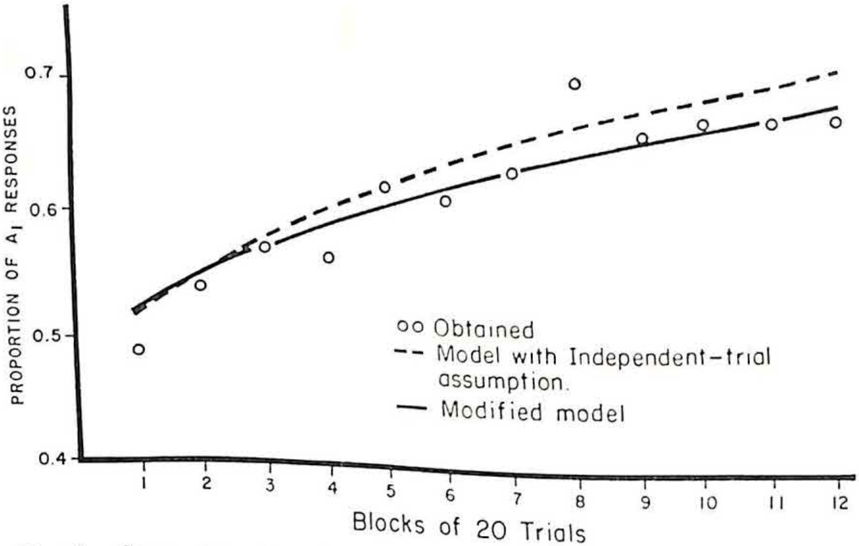


FIG. 2. Comparison of predicted proportions of A_1 responses for two models with obtained data for Group IV.

In the constant-signal group (Group VII), θ_s is larger than θ_p, but the values for α and θ are practically identical. In fact the theoretical curves predicted by both models are identical for this group from the fifth to the last block of trials. It is only in the variable-signal groups that the values for α and θ are disparate, and in these cases the modified model provides the better fit. Figure 2 presents the empirical acquisition curve for Group IV, together with the theoretical curves predicted by both models. Group IV is chosen, in that with $\theta_p^* > \theta_s^*$, it involves the most variable signal of the experimental groups.

The relationship between α and θ was further explored by the following procedure. The mean acquisition curve was estimated for each of the seven groups—using the independent learning model equation, which is

$$p_n = \pi - [\pi - p_1](1 - \theta)^{n-1} \qquad [7]$$

but the obtained value of α, given in Table 2 was substituted for the learning parameter θ in Equation 7. For Group VII the curve was the same whichever method we used—which as $\alpha = \theta$ in this case, is not surprising. But also, for every group with variable signal, the curve obtained by this method corresponded almost exactly to those generated by θ_s and θ_p of the modified model. Apparently the ultimate proportion of elements in the sample is represented adequately by the formulation for α, but θ of the independent sampling model is not always identical with that proportion.

A final interesting point about the α values is that it is possible to derive from the present modified model, precise predictions of the comparative magnitude of the slope parameters for the seven groups. The derivations although simple, are involved, and will not be presented here. However, the predicted order of magnitudes is identical with the ordering of the α values shown in Table 2.

And finally, the basic hypothesis that in a massed-trial situation $\theta_s > \theta_p$ is very well supported by the values of the parameters listed in Table 2. For every group—including the constant-signal group—the value of θ_s is greater than that of θ_p. Moreover for the two experimental variable-signal groups (Groups II and III) where the signal parameters showed a similar relationship; i.e., $\theta_s^* > \theta_p^*$, the ratio of the estimated θ_s/θ_p is considerably greater than for any of the other variable-signal groups. The a priori expectation that the experimental manipulation of the signal would amplify the hypothesized "massed-trial" effect in the case of these groups was therefore confirmed.

DISCUSSION

For the group where the signal does not vary, both the "independent" sampling model and its present modification are equally successful in predicting the mean curve. Inspection of the values listed in Table 2 indicates why this is so. Both α, the proportion of elements in the stimulus sample, and θ_s, the overlap parameter, are very small. The proportion of elements remaining in the sample should therefore have little effect upon the mean learning curve. Assumption of trial independence therefore, would not be expected to interfere with adequate mean learning-curve predictions in the usual massed-trial situation where the signal remains, as it usually does, constant from trial to trial.

The relationship between α and the independent-model parameter θ is an interesting one. Both are intended to represent the proportion of elements in the sample, and insofar as the group with constant signal is concerned, the values for both parameters are identical and describe the slope of the learning curve adequately. When the signal varies in the rather complex manner described here, α still describes the learning curve slope efficiently, θ does not.

Probably the most pertinent findings of the present study for statistical learning theory is that the disparate values obtained when θ is estimated from two sources which should yield the same value—a problem described earlier—would be expected from the point of view of the modified model. In this model the difference between conditional probabilities represents the overlap parameters θ_s. In a massed-trial situation θ_s is shown (Table 2) to be greater in both variable and constant-signal groups than the values of θ_p and α. But according to our present model α is analogous to the independent sampling model's single parameter θ, and where the signal remains constant, would be expected to yield the same value. Consequently we would expect that the difference between the conditional probabilities would yield the larger value.

REFERENCES

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 381–491.

GINSBERG, R. Correlated stimulus samples in verbal conditioning. Unpublished doctoral dissertation, Indiana University, 1959.

STRAUGHAN, J. H. Human escape learning in relation to reinforcement variables and intertrial conditions. *J. exp. Psychol.*, 1956, **52**, 1–8.

(Received July 15, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 560–572

SOME CONDITIONS SUFFICIENT FOR ACCURATE MONOCULAR PERCEPTIONS OF MOVING SURFACE SLANTS[1]

HOWARD R. FLOCK

Hunter College of the City University of New York

246 Ss monocularly judged from 1 to 9 different slants of a surface whose edges were occluded from view. In a reference experiment surface texture was irregular, unfamiliar, and continuously moving, viewed through a 69° aperture. Then the mean regression coefficient for 16 Ss' slant judgments over 9 physical slants was .96, a coefficient of 1.0 signifying an identity relation. Doubling and halving velocity, substituting regular or different irregular textures, changing amount of exposure to the displays did not significantly alter the results—the coefficients varying from .85 to 1.12. Absence of motion, a texture of ½-in. squares, and changing field of view over a series of 4 physical slants alone produced different results. Dependence of slant on shape judgments was also tested. Shape judgments varied positively with retinal shapes rather than with slant judgments. Individual data were evaluated.

In a series of theoretical articles Gibson (1950, 1961; Gibson & Gibson, 1955, 1957; Gibson, Olum, & Rosenblatt, 1955) has argued that the light reflected to the single eye from a surface necessarily carries sufficient information about the spatial properties of the surface. Gibson and his colleagues (Flock, 1962; Gibson & Gibson, 1957; Gibson et al., 1955; Purdy, 1958) have supported these arguments with mathematical analyses of the array of light reflected from surfaces to the eye.

Flock (1962), for example, employing trigonometry, isolated a ratio of ratios of angular motions in the light which was reflected to the eye from a slanted and moving surface. It was shown that those angular relations accurately specify the slant of the surface regardless of its velocity and even though the surface's texture is novel, the texture elements irregular in size and shape and irregularly distributed across the surface, and the surface indefinitely extended (as when a surface's edges are occluded from view). On the basis of that analysis of the light there was at least a theoretical prediction that variables of velocity and texture would not prove to be important factors in the accurate monocular perception of the slant of a moving surface.

Nevertheless, while general theoretical predictions like this one have been made or implied by Gibson and his colleagues, little systematic attempt has been made to test their truth and under what conditions they might be true. This paper, therefore, will attempt to investigate some of the conditions under which the theoretical prediction above might be true. Stimulus conditions will first be described which elicit relatively accurate judgments of nine different slants of a presumptively unfamiliar, irregularly textured, moving surface. Those conditions and the resulting level of judgmental accuracy will constitute a reference experiment. Attempts will then be made to replicate the results of the reference experiment under different conditions which according to the predictive hypothesis should not significantly alter the judgmental results. Tests will be made of the effect of changing velocity, of substituting a different irregular texture or three regular textures, of changing the amount of exposure to the reference situation, of reducing velocity to zero, and of changing the size of the field of view. Finally, since the predictive hypothesis precluded any dependence of slant on shape judgments, an experiment will be described evaluating the relation between slant judgments and the judgment of at least one shape amongst the textural elements on a surface.

---

[1] This research was supported in part by the Office of Naval Research under contract NONR 401 (14) with Cornell University; and by a Public Health Service Research Fellowship from the Department of Health, Education, and Welfare. It is based on a thesis submitted in partial fulfilment of the requirements for the PhD degree, Cornell University, 1962, supervised by J. J. Gibson to whom the writer wishes to express his gratitude.

EXPERIMENTS I–X

Method

Apparatus.—In order to prevent primary perceptual depth cues from influencing the results Gibson and Gibson's (1957) point-source shadow projector was employed. This optical device simulates at the eye those luminous relations which are found when a slanted surface is distally present. Briefly, the point-source shadow projector places a textured and slanted surface in a projection system so that *S* monocularly views a projection of the textured surface to a frontal plane instead of viewing the slanted surface itself. The system contains five major elements: a point source of light; a shadow caster (transparent medium textured by opaque elements distributed across it); a mount for holding the shadow caster at some slant and/or moving it in a certain way; a translucent projection screen; and a station point located such that a line joining the point source and station point is perpendicular to the projection screen and is bisected by the projection screen.

The arrangement of the apparatus, as seen from the side, is shown in Fig. 1. On one side of the projection screen (a 6 × 6 ft. sheet of .188-in. white translucent Plexiglas, Rohm and Haas No. 267) was placed the shadow caster in its mount, its horizontal axis being 6.72 in. (16.8 cm.) from the point source and 29.68 in. (74.2 cm.) from the screen. Behind the shadow caster, 36.4 in. (91 cm.) from the screen, was the point source of light (a 25-w. Sylvania concentrated-arc end light with a light element that was .03 in. in diameter, powered by a 243-dc power supply). On the opposite side of the screen, 36.4 in. (91 cm.) away, was the station point at which *S* placed his eye, viewing the display on the projection screen through a fixed 69° round opening in an iris diaphragm. The projection screen bisected and was perpendicular to the axis running through the station point and the light element.

The shadow caster mount was constructed of wood and functioned to hold the shadow caster at any one of nine different rotations on its horizontal axis: 40°, 30°, 20°, 10°, 0° (frontal position), −10°, −20°, −30°, −40°. At the same time that the surface was slanted it could be moved back and forth so that its horizontal axis was always the same distance from the projection screen.

The *S* was separated from the apparatus by a black curtain wall in the middle of which was an aperture with a fitted eyepiece. Combined with the aperture was an Ilex No. 5 shutter with a cable release that allowed *S* to open and shut the aperture. To the right of *S* at his eye level and facing him when he turned towards it was an 11-in. disc divided into a black and white half. The diameter specified by the two colored halves could be rotated in a clockwise or counterclockwise direction to any degree of rotation so that the direction of rotation matched the angle of slant of the moving surface. This apparatus for measuring the slant judgments is hereafter referred to as a protractor. By use of a synchronous motor arrangement *S*'s protractor drove a comparable disc graduated into degrees from which *E* seated behind the apparatus could read off the protractor adjustment.

Shadow casters.—In the experiments three types of surface texture were used as shadow casters. They were made by placing opaque material on 36 × 36 in. sheets of transparent Plexiglas, .188 in. thick. One type of surface texture was composed of a grid of squares, the squares being either $\frac{1}{2}$ in., $\frac{1}{4}$ in., or $\frac{1}{8}$ in. along an edge. These surfaces were called regular textures. The $\frac{1}{2}$-in. Regular Texture was made by laying $\frac{1}{2}$-in. strips of black opaque chart tape in parallel bands separated by $\frac{1}{2}$-in. strips of blank Plexiglas. The surface

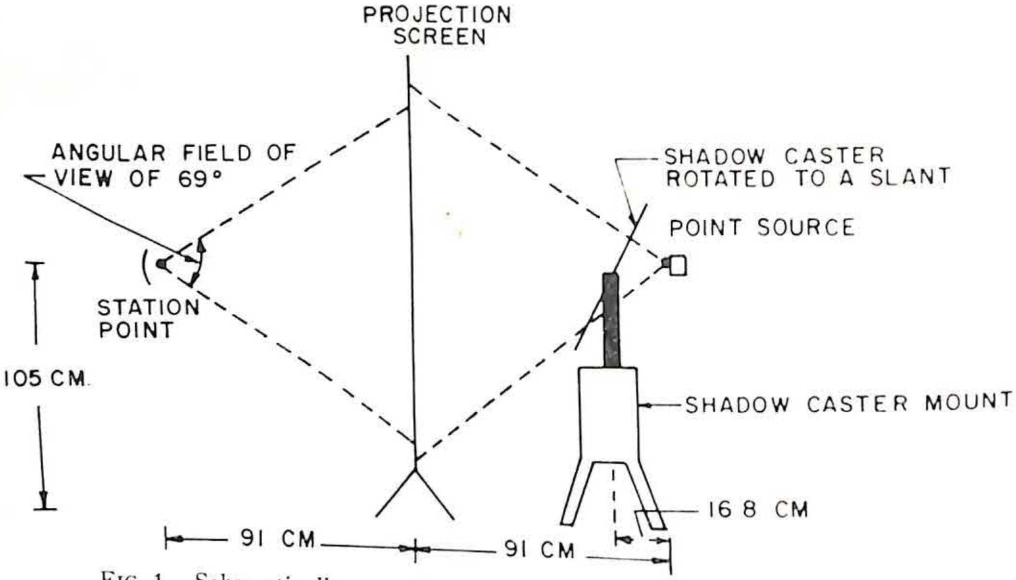


FIG. 1. Schematic diagram of the point-source shadow projector: side view.

was then rotated 90° and the procedure repeated. The $\frac{1}{4}$-in. and $\frac{1}{8}$-in. Regular Textures were made in the same way, the separations and bands then being $\frac{1}{4}$ in. and $\frac{1}{8}$ in., respectively, instead of $\frac{1}{2}$ in.

The second class of surface textures was composed of a highly dense array of randomly irregular, randomly achromatic shapes distributed irregularly across the transparent Plexiglas surface. These surfaces, hereafter referred to as electrostatic surface textures, were made in a completely novel manner (see Flock, 1964a). When a Plexiglas surface is given an electrostatic charge and if a pen is at the same time held above the surface, ink will flow to and spread over the surface in a pattern dictated by the charges on the surface. Two of these electrostatic surfaces were constructed. The first, Electrostatic Texture A, differed considerably from the second, Electrostatic Texture B. A third type of irregular surface called a Powder-Sprinkle Texture was made by sprinkling talcum powder on the Plexiglas surface. (The powder-sprinkle and electrostatic surfaces are described and illustrated in Flock, 1964a.)

Helmholtz (1925, pp. 283, 285) described textures like the regular textures, pointing out that they are often experienced by and therefore are familiar to adults in our culture. He also described what he meant by unfamiliar textures, indicating, for example, pictures of rock piles or a snow field in which the texture elements are irregular in shape and distribution (Helmholtz, 1925, p. 285). Since the electrostatic textures were not only highly irregular and made in a novel way but also did not seem to correspond to anything commonly experienced in the physical environment, it seemed that they more than satisfied the criteria set forth by Helmholtz for an unfamiliar surface texture.

Procedure.—Each *S* was told that he would see a surface which would be slanted in some way, perhaps with its top towards him and the bottom away, or with its top away from him and bottom towards, or from top to bottom in a frontal position. Each *S* was given detailed instructions in the use of the protractor, rotating the diameter in one direction if he saw a surface with its top slanted away from him, setting the diameter vertical if he saw a frontal surface, etc. He was particularly told to rotate the diameter so that its angle of rotation exactly matched the surface slant he saw. No time limit on the judgment was imposed. The *S* was instructed to examine his entire field of view, closing the aperture only after he was satisfied with the protractor adjustment.

Subjects.—The 246 *S*s, 82% of whom were men, were for the most part students from the introductory classes in psychology at Cornell University. Their ages ranged from 17–27, with a median age of 18.5. Since the eyepiece would not accommodate eyeglasses, no *S* of the 54% who ordinarily wore glasses was allowed to wear glasses during the experimental sessions unless he happened to be wearing contact lenses. All *S*s were volunteers and naive to the purpose of the experiments.

Experimental design.—The 246 *Ss* were assigned to 10 subgroups, 7 of the groups each having 16 *Ss*, 2 of the groups each having 55 *Ss*, and 1 having 24 *Ss*. Each subgroup was assigned to a particular condition of textured surface, either to a regular, familiar surface or to an irregular, unfamiliar surface. For all *Ss* in 9 of the groups the textured surface was in a continuous, side-to-side lateral motion, the motion being at right angles to the axis through the station point and the light element. At no time was the edge of any surface projected onto the screen. Nor was any *S* ever able to see the edges of the projection screen. Except for *Ss* of one group the field of view was always 69°.

In Table 1 the various experimental conditions are listed. The conditions cited for Exp. I (Table 1, Row 1) constitute the reference conditions in comparison to which all of the remaining experiments were replications. For example, Exp. II–III differed from Exp. I only in the velocity of the shadow caster. Experiment IV was the same as Exp. I except for the change in the irregular texture. Experiments V–VII were various replications of the effects of three different regular textures on slant judgments. Each *S* assigned to Exp. I–VII made protractor judgments first of the nine different slants (40°, 30°, 20°, 10°, 0° [frontal position], −10°, −20°, −30°, −40°) of the surface, presented in random order, and second of the same slants in a different random order, a total of 18 judgments for each *S*. Each pair of judgments for each level of slant was then averaged, yielding 9 mean judgments for each of the *Ss* of Exp. I–VII.

In Exp. VIII all conditions but two were the same as in Exp. I. The group of 55 *Ss* was subdivided into five equal subgroups, each subgroup being assigned to a different slant of the surface. Whereas *Ss* of Exp. I each made 18 observations, 2 each for nine different slants, *Ss* of Exp. VIII each made but one judgment. In addition, a small red translucent ellipse, 19.2 × 15 mm. (the purpose of which will be described later) was inked onto the center of and embedded amongst the shapes of Electrostatic Texture A. In Exp. IX, except for a change in the velocity of the shadow caster to zero, all conditions were the same as for Exp. VIII. Since both the design and procedure for Exp. X differed from that for Exp. I–IX, no further reference will be made to Exp. X until its results are later described. A final experiment, Exp. XI, will also be described at a later place. The chronological order in which the experiments occurred could conceivably have confounded the results. Some of the experiments were run concurrently, others were not.

In order to evaluate the effects of the various conditions on *Ss*' judgments, each *S*'s nine mean judgments for Exp. I–VII were reduced to a linear regression coefficient. The distribution of these coefficients under each experimental condition gave an estimate of the intervariability amongst *Ss* and was used to test for any differences between the mean coefficients for the various experiments. The model for this treatment of the data was based on Grant (1956). The statistical models for the other experiments will be described later.

Results and Discussion

In Table 2 is listed the frequency of occurrence of specific regression coefficients under Exp. I–VII. Ex-

TABLE 1

VARIABLES OF TEXTURE, VELOCITY, ANGULAR FIELD OF VIEW, AND LEVELS OF PHYSICAL SLANT FOR 10 EXPERIMENTS

Exp.	Surface Velocity in Ft/Sec	Field of View	Composition of Surface Texture	*N*	Levels of Physical Slant/*S*
I	1/8.3	69°	Electrostatic A	16	9
II	1/3.81	69	Electrostatic A	16	9
III	1/14.54	69	Electrostatic A	16	9
IV	1/8.3	69	Electrostatic B	16	9
V	1/8.3	69	$\frac{1}{8}$-in. squares	16	9
VI	1/8.3	69	$\frac{1}{4}$-in. squares	16	9
VII	1/8.3	69	$\frac{1}{2}$-in. squares	16	9
VIII	1/8.3	69	Electrostatic A	55	1
IX	0	69	Electrostatic A	55	1
X	1/8.3	17, 35, 52, and 69	Powder-Sprinkle	24	4

TABLE 2

Frequency Distribution of Regression Coefficients for Exp. I–VII, Means and Sigmas

Regression Coefficients	Experiments							Row Totals
	I	II	III	IV	V	VI	VII	
1.80–1.91			1					1
1.68–1.79		1						1
1.56–1.67								0
1.44–1.55				2				2
1.32–1.43			1					1
1.20–1.31	2	2	1	1		2		8
1.08–1.19	2	2	2	2	1	2		11
.96–1.07	4	3	4	4	4	3	2	24
.84– .95	4	6	5	2	4	4	4	29
.72– .83	2		1	3	3	1	6	16
.60– .71	2	2	1	1	2	4	3	15
.48– .59				1	2		1	4
M	.96	1.02	1.05	.99	.85	.91	.79	.94
s	.166	.248	.274	.24	.17	.20	.16	.225
N	16	16	16	16	16	16	16	112

Note.—Means and sigmas are based on ungrouped data; *CE* of each mean is equal to: $M - 1.0$; $s_b = s/\sqrt{N}$.

amination of the table indicates that no *S*'s coefficient was under .48 nor greater than 1.91. Except for two *S*s distributed between Exp. II and III the total range would have been from .48 to 1.55. Had there been an identity relationship between judged and physical slants, the observed coefficients would have been 1.0. Hence a coefficient of 1.0 constituted a criterion of maximally possible accuracy.

From the scatter of the individual coefficients under each experiment there is the suggestion that except for a slight positive skew the distribution of the coefficients might approach normality. Further, *F* tests of the variances (inter-*S* variability of coefficients) from one experimental condition to a second led to the conclusion that the variances were homogeneous. (The model for these tests was based on Ryan, 1960.)

Reversals.—The use of linear regression coefficients as a measure of accuracy of *S*s' judgments omitted any reference to possibly significant, nonlinear fluctuations in *S*'s data. As a rough measure of the degree of such nonlinear effects the reversals in each *S*'s data were counted. Consider, for example, a sequence of three physical slants, 5°, 10°, 15°. If *S* reported these to be 8°, 7°, and 15°, respectively, there would be 1 reversal. Had he reported them as 8°, 7°, and 6°, respectively, there would be 3 reversals. Correspondingly, for a sequence of nine physical slants, 36 reversals would be possible and 18 might be expected to occur by chance. In an experiment with 16 *S*s, 576 reversals would be possible. For Exp. I–VII, the reversals were 1, 10, 8, 5, 1, 7, 2, respectively, a total of 34 reversals out of 2,016 expected by chance. Moreover, 87 of the 112 *S*s had no reversals, of which 9 *S*s had 1 tie; 17 *S*s had 1 reversal; 7 *S*s had 2 reversals; and 1 *S* had 3. Of the 34 total reversals 20 would have disappeared with a change in slant judgments of 2.5° or less; and 13 more would have disappeared with changes of less than 9°.

It seemed, therefore, that the reversals were sufficiently rare to warrant the use of individual linear regression coefficients. This use of linear regression coefficients might have ignored nonlinear effects in the individual data. But on the reasonable assumption that an S's error variance (unmeasured in these experiments) was something greater than zero such nonlinear effects, were they to exist, would have been moderate.

Reference experiment.—In Table 2 Column 1, the mean coefficient for the reference experiment, Exp. I, is .96 ($s = .17$). The mean result varied from the criterion of maximal accuracy ($b = 1.0$) by an insignificant error of $-.04$. As specified in Table 1, the surface texture consisted of irregular shapes irregularly distributed, and the velocity of the shadow caster was 1 ft/8.3 sec.

Effects of velocity changes and Electrostatic Texture B.—When velocity of the shadow caster was doubled and halved as in Exp. II and III, respectively, the mean coefficients were 1.02 and 1.05, respectively. Multiple comparisons between each of these means and the mean of .96 from Exp. I yielded insignificant t scores of .76 (30) and 1.05 (30), respectively; $\alpha = .05$ adjusted for six tests.[2] That the results of Exp. I–III were not peculiar to Electrostatic Texture A may be concluded from contrasting Exp. IV in which Electrostatic B replaced Electrostatic A with Exp. I, the two means being .99 and .97, respectively. The t test between the means yielded an insignificant $t\ (30) = .34$.

Practice effects.—If S somehow learned about the surface shapes during the 18 trials of the experiment, and if this learning accounted for the fact that the mean regression coefficients were both positive and near 1.0, then one might expect that if Ss were given but one trial the regression coefficient would be at or near zero.

To test this possibility each of the five subgroups in Exp. VIII was assigned to one of the following levels of physical slant: 40°, 20°, 0° (frontal), $-20°$, $-40°$. Thus each of the 11 Ss in a subgroup made just one judgment of slant. Except for the presence of the small red ellipse in the center of Electrostatic A all other conditions were like those in Exp. I. A regression line was fitted to the judged mean slants over the five levels of physical slant, giving a mean coefficient of 1.12. From the pooled within-group variation, the standard error of the slope was .05.

While a direct statistical test could not be made between Exp. VIII and I owing to the incommensurate nature of the error terms, a comparison can be made between the two experiments. With only a single trial in Exp. VIII, the *CE* of .12 was a little larger than the *CE* of .04 in Exp. I where Ss had 18 trials. The difference of .08 between *CE*s hardly suggests that the absence of prior experience with the irregular, unfamiliar textures constituted a major factor in the accuracy of the judgments.

Effect of regular vs. irregular textures.—In Exp. V–VII the mean regression coefficients were .85, .91, and .79 for the $\frac{1}{8}$-in., $\frac{1}{4}$-in., and $\frac{1}{2}$-in. textures, respectively. When these means were compared with the mean from Exp. I, the t's (30) were 1.8, .8, and 3.1, respectively. Of these only the difference between the means for the ½-in. and Electrostatic A textures was significant ($p < .008$). Nevertheless, when the degree of accuracy represented by the results for the regular and ostensibly familiar textures (Exp. V–VII) was compared with the results for the irregular and unfamiliar electrostatic surfaces (Exp. I–IV and VIII) the *CE*s for the regular textures were uniformly greater than those for the irregular textures.

---

[2] The significance level employed throughout all experiments was $\alpha = .05$. When multiple t tests were made, however, alpha was adjusted to maintain the .05 level throughout a group of multiple comparisons of the same kind (see Ryan, 1959, 1960). Since the means from Exp. II–VII were each compared to the mean from Exp. I—a total of six t tests—alpha was adjusted for these tests to $.05/6 = .008$. Therefore, significance in these tests is given by $p < .008$, although $\alpha = .05$.

This phenomenon, that irregular and novel textures could have yielded more accurate results than did regular and familiar textures, might have been an artifact of one particular property of the textures used. The phenomenon might be accounted for by the differences in density amongst the surface textures. In comparison with the electrostatic surfaces the regular textures had a relatively small number of textural elements for the fixed 69° field of view. Moreover, in the data for the regular textures alone, there was at least a suggestion that density could have been a factor. The mean coefficient for the ½-in. texture was .79, the means for the more dense ¼-in. and ⅛-in. textures being somewhat closer to 1.0. On the other hand, the slight trend in the data supporting density as a factor may turn out to be adventitious. Certainly more research on this problem is indicated.

Zero velocity.—The predictive hypothesis governing the experimental findings required that a surface be in motion. In Exp. IX conditions were exactly the same as in Exp. VIII, the 55 *S*s of Exp. IX each making a single slant judgment, but in this instance the shadow caster was motionless. The coefficient through the five group means for Exp. IX was then .13, with $s_b = .10$; whereas it was 1.12, with $s_b = .05$, for Exp. VIII. The difference between these means was significant, $t\ (100) = 8.7,\ p < .05$.

Effect of field of view.—In Exp. X a group of 24 *S*s twice judged four slants (30°, 20°, 10°, 0°) of a surface under each of four different fields of view (17°, 35°, 52°, 69°), making a total of 32 observations for each *S*. Except for the use of a powder-sprinkle texture all other conditions were the same as in Exp. I.

Regression coefficients were calculated for *S*s' judgments of the four slants under each level of field of view. Those were: .82 for 17°; 1.09 for 35°; 1.22 for 52°; and 1.15 for 69°. The only significant differences between pairs of these were between the smallest and the two largest coefficients. But on the basis of the constant errors it would seem that the slant judgments were as accurate for the 17° as for the 52° and 69° fields of view.

The conditions in this experiment differed from those in Exp. I in three major respects. (*a*) The *S*s were exposed to the displays for a longer time (32 instead of 18 judgments) under varying conditions (changing field of view and slant). (*b*) The *S*s viewed four instead of nine slants. (*c*) A powder-sprinkle rather than an electrostatic texture was used. Any one or several of these differences could have produced the large constant error for the 69° (.15) and 52° (.22) fields of view. In view of the findings reported above under comparable conditions, however, it may be that the main factor producing the large errors was the use of a series of four rather than nine slant treatments. The raw data tended to support this conclusion. For the 69° and 52° fields of view the mean judged slants were: 32.4°, 22.2°, 10.8°, −2.3°; and 33.2°, 19.5°, 9.2°, −4°, respectively. Both series have two common features: a tendency for the ends of the series to pull away from the middle and a tendency for the judgments in the middle of the series to be approximately accurate.

Regardless of the reason for the end effects, they produced the large constant errors. Moreover, in a series with nine slant treatments the same end effects would have less influence on the regression coefficients. For example, if one applied the same constant errors for the 30° and 0° slants to the 40° and −40° slants in a series of nine slants, the regression coefficients would be 1.03 and 1.05 instead of 1.15 and 1.22 for the 69° and 52° fields of view. It may be, therefore, that the mean coefficients for the four fields of view were all too high by approximately .12–.15 points of slope and should be adjusted accordingly. Certainly the effect of changing field of view as well as the effect of changing number of slant-treatment levels require further investigation. (The details of Exp. X and the analysis of the data are reported in Flock, 1962.)

EXPERIMENT XI

Flock's (1962) trigonometric analysis of the optical stimulus for slant precluded any statement about the shapes of the elements composing a surface's texture. On that basis, therefore, regardless of the conditions of stimulation, there is no theoretical necessity for arguing that slant judgments will covary with shape judgments. As a corollary to that, when conditions exist for and at the same time induce accurate slant judgments, there is no theoretical necessity to expect that slant and shape judgments will even then covary. These hypotheses differ from Koffka's (1935) assertion that slant and shape judgments must covary and differ from Beck and Gibson's (1955) expectation that while they need not covary when conditions for accurate slant judgments are minimal, they should tend to covary when conditions for accurate slant judgments are relatively optimal. The experiment described below tested these various hypotheses regarding the presence or absence of what has been called a shape-slant invariant. For a general review of the past methods employed and attempts to test the shape-slant invariant hypothesis, see Epstein and Park (1963).

Method

Apparatus, procedure, and Ss.—Each group of 55 *S*s in Exp. VIII and IX was required, following their slant judgments, to attend the projection of the 19.2 × 15 mm. red ellipse in the center of the display. The ellipse was colored so as to be identifiable. When the surface was motionless the ellipse lay approximately astride the axis of rotation of the surface and was projected to the middle of the display on the projection screen. When the surface was in motion the projected ellipse moved back and forth continuously across the projection screen, lying in the axis of rotation of the shadow caster.

The projected ratios of height to width of the ellipse on the projection screen for the five slants of the shadow caster were: .95:1 at 40°; 1.20:1 at 20°; 1.28:1 at 0° (which is the actual ratio of the height to width of the ellipse in the plane of the surface); 1.22:1 at −20°; and 1.01:1 at −40°. The projected ratios for the pair of 40° and of 20° slants were slightly different owing to the ellipse being slightly off the axis of rotation of the surface. At the 40° slants the ellipse projected a retinal shape that was approximately a circle.

After each *S* had made his slant judgment he re-examined the display and made a judgment of the shape proportions of the embedded ellipse first under an "appearances" instruction and then under an "objective" instruction. Each of the 110 *S*s thereby made one slant judgment and two shape judgments.

Under the appearances instruction each *S* was told that he would see a red figure in the middle of his field of view. "If you assume that this red figure is 2 in. wide, at its widest place (which was along its horizontal axis), then how high in the up-down direction does the figure appear to be?" (This judgment of the height in inches is later extrapolated to a ratio of height to width, with width taken as unity.) After *S* had given the first shape judgment, he was given slightly more objective instructions and asked to make a second shape judgment. "Look again at the red figure. If you could go up to the surface and make a tracing of the figure by laying a piece of tissue paper on it, then how long and how wide would it appear to be?"

There were obvious differences between the two sets of instructions. In the first set *S* was told to assume that the object was 2 in. wide, and on the basis of that information to estimate its height. The second set of instructions allowed *S* to express the height and width dimensions in any metric that suited him.

The second difference in the instructions was more important. In the first set (the appearances instruction) *S* was asked how high the figure *appeared* to be. In contrast the second set (the objective instructions) emphasized the objectivity of tracings and of things as they really are. But then the objective instructions ended by asking for the appearance of the shape. In this way *E* hoped to prevent *S*s from *betting* what the shape dimensions were and thereby perhaps introducing idiosyncratic cognitive components into the judgment, as might have happened in Stavrianos' (1945) experiment.

The total time of an experimental session, excluding instructions about the use of the protractor, was about 2.5–3 min. per *S*.

Experimental design.—Each *S*'s slant judgment was converted into an expected shape judgment. For example, consider three *S*s under the 20° moving slant condition who might have judged the surface to be slanted 30°, 20°, and 10°, respectively. For the constant retinal proportions of the projected ellipse when at 20° slant, if *S* incorrectly sees the object as slanted to 30°, then under the hypothesis of shape-slant invariance he should report a shape greater than the 1.28:1 proportions of the objective ellipse. These expected shape judgments can be calculated from the following ratio: (retinal shape proportions of the ellipse at a particular slant)/(cosine of the judged slant for that condition). For the three hypothetical *S*s above the three expected shape judgments would then be: 1.39:1, 1.28:1, and 1.22:1, respectively. As would be expected, if the actual slant is 20° and it is judged as 20°, then the expected shape judgment would be 1.28:1, which are the actual shape proportions of the objective ellipse.

For each of the 110 *S*s, therefore, there were one expected shape judgment based on his slant judgment and two observed shape judgments, one each under two instruction conditions. If there is a necessary invariance between shape and slant judgments, then the regression of observed on expected shape within each slant-treatment group should be homogeneous over the 10 treatment groups (2 motion × 5 slant groups) under an instruction condition. Moreover, if homogeneity of the within coefficients exists, pooling them should produce a coefficient that is both positive and significantly different from zero.

A second test of invariance, disregarding the presence or absence of positive covariation within the treatment groups, may be made by describing the relationship between the observed and expected mean shape judgments for the four Motion × Instruction conditions and by comparing the mean regression coefficients describing the relationship.

Results and Discussion

The test for the homogeneity of the 10 coefficients within the slant-treatment groups under the two instruction conditions is described in Table 3. The third *MS* in the table (Row 3) is derived from the deviations within the 10 treatment groups from a pooled regression for all within-treatment groups. The fourth *MS* (Row 4) is based on the deviations within each of the 10 groups from the regression within the group. The fifth *MS* is calculated by subtracting the *SS* for Row 4 from *SS* for Row 3. The *F* ratios of .33 (9, 90) and .2 (9, 90) were not significant, $p > .025$. From this evidence it seems, therefore, that the within coefficients are not heterogeneous, neither under the appearances nor under the objective instructions. But this finding of homogeneity does not imply invariance of shape and slant judgments. The pooled within coefficients were −.08 and −.11 (Table 3, Row 6) under the appearances and objective instructions, respectively. Not only were those two regression coefficients not significantly different from zero, *F*s (1, 99) = .71 and .63, $p > .025$, re-

TABLE 3

ANALYSIS OF THE HOMOGENEITY OF REGRESSION OF JUDGED ON EXPECTED SHAPE WITHIN 10 TREATMENT GROUPS FOR EACH OF TWO INSTRUCTION CONDITIONS AND THE SIGNIFICANCE OF THE POOLED REGRESSIONS

Source	Instructions				
	Appearances			Objective	
	df	*MS*	*F*	*MS*	*F*
Total variance within	100	.084	—	.144	—
Pooled regression (RG) within	1	.06	.71	.091	.63
Error of estimate within	99	.084	—	.145	—
Deviations from a group's RG	90	.09	—	.154	—
Diffs. among group RGs	9	.03	.33	.038	.2
Coefficient for pooled within-treatment-group RGs		−.08		−.106	

Note.—These tests are described by Snedecor, 1948, pp. 325 ff. and especially Tables 12.7 and 12.5.

spectively, but their signs were not positive as would be required under a hypothesis of invariance. Moreover, of the 20 within coefficients, 10 each for the two instruction conditions, 12 of them were negative, 5 were zero, and only 3 were positive. In short, of the 20 coefficients only 3 showed any tendency at all to support the hypothesis of shape-slant invariance.

This evidence seemed to give no support to Koffka's (1935) hypothesis that regardless of the conditions of stimulation there will be a relation between slant and shape judgments. Nor was there evidence here for Beck and Gibson's (1955) expectation that the judgments should tend to covary if the conditions for slant judgments induce relatively accurate slant judgments. Rather, it seemed that in this experiment the relationship between slant and shape judgments was a chance relationship throughout.

If one disregarded the absence of relationship within the treatment groups, was there at least a tendency for the observed mean shape judgments to covary positively with the expected mean shape judgments? In Table 4, Row 2, are given the coefficients of regression of observed mean shape on expected mean shape judgments under the two instruction conditions and the two motion conditions. Under the appearances instruction the coefficients were −1.17 (motion) and +1.66 (no motion); and under the objective instruction they were −.49 (motion) and +1.21 (no motion). All of these coefficients were significantly different from zero, their *t*'s (99) being 12, 17, 3.8, and 9, respectively, $p < .012$, the error terms having been derived from Table 3, Row 3.

Note that the two negative coefficients under the motion condition contradict the invariant hypothesis. In contrast, under the motionless condition the coefficients are positive

TABLE 4

COEFFICIENTS OF REGRESSION OF MEAN JUDGMENTS (JS) OVER FIVE TREATMENT LEVELS UNDER TWO VELOCITY AND TWO INSTRUCTION CONDITIONS

Relationship	Motion		Motionless	
	App.	Obj.	App.	Obj.
Slant Js on physical slants	1.12		.13	
Shape (SH) Js on expected (EX) SHs	−1.17	−.49	1.66	1.21
EX SH Js on retinal SHs	−.03		.97	
SH Js on retinal SHs	1.17	1.34	1.65	1.20

and significantly different from zero. But these significant positive regressions under the motionless conditions do not imply a relation of invariance between the observed and expected shapes. In Row 3 of Table 4 is given the coefficient of regression of expected mean shapes on retinal shapes under the motionless condition. That coefficient is +.97 and is significantly different from zero, t (50) = 19.4, $p < .025$, s_b being .05. From this coefficient of +.97 one may conclude that a relationship existed between the expected mean shape judgments and the retinal shapes. But if the expected shapes covaried positively with the retinal shapes, then the significant, positive covariation between observed and expected shapes that occurred in the data for the motionless condition implies that the observed shapes covaried positively with the retinal shapes. (In contrast notice that under the motion condition the coefficient of regression of expected on retinal shapes was approximately zero, −.03, Table 4, Row 3.) The possibility that the shape judgments in this experiment covaried positively with the retinal shapes rather than with the slant judgments is next considered.

When regression coefficients were computed between the four sets of five observed mean shape judgments and the five perspective shapes, all of the four regression coefficients (Table 4, Row 4) were positive and significantly different from zero, the t's (100) being 13, 12, 19, 11, respectively, $p < .012$. The tests of significance were made with error terms computed from Table 3, Row 1. This finding implies that the mean shape judgments regardless of the presence or absence of motion, or of the change in instructions varied with the retinal shapes, not with the slant judgments as required by the various invariant hypotheses described earlier.

Beck and Gibson (1955) had found that shape judgments may vary at random when conditions do not support accurate slant judgments. Under those conditions, they asserted, the shape judgments need not vary with the retinal shapes. That prediction seems not to hold in this experiment.

Brunswik (1944, 1956) had emphasized the effects of instructions on judgments, expecting different instructions to produce radically different judgments. The appearances instruction under the motion condition failed to elicit accurate shape judgments, as evidenced by the coefficient (1.17) of regression of mean shape judgments on perspective retinal shapes. Presumably the appearances instructions induced what Brunswik (1944, p. 4) referred to as an "analytical perceptual attitude" (proximal focusing). Under this attitude *S*s presumably analyze the display and then "rely exclusively on perceptual appearances" in judging the shape. To induce correct shape judgments (distal focusing) Brunswik (1956, p. 22) recommended instructions which "drastically emphasized" the object qualities. Since the objective instructions in Exp. XI possessed that emphasis, Brunswik would have expected them to induce good shape judgments, at least under the motion condition, and therefore a positive covariation between mean shape and slant judgments. That different effect of the objective instructions did not occur. The mean shape judgments under the objective instructions varied with the retinal shapes under the motion condition ($b = 1.34$) and under the motionless condition ($b = 1.20$). Moreover, the difference between the effects of the two different instructions under the motion condition (1.34 vs. 1.17) was not significant. While the difference between the effects of the two instructions under the motionless condition (1.65 vs. 1.20) was significantly different, the direction of the difference fails to support Brunswik. The objective instructions were observed to produce a coefficient (1.20) closer to 1.0 than was the case for the appearances instructions (1.65). Following Brunswik, one would expect the appearances instructions rather than the objective instructions to produce a coefficient closer to 1.0.

Finally, two procedural variations in this experiment require amplification. The *S*s had been asked to judge the slant of a surface composed of many shapes, but to judge the height-width ratio of only one elliptical shape embedded in the surface. This procedure seems to be warranted by prevailing theories of space perception. In those theories the registration at some level of accuracy of the slant of a surface seems to constitute a sufficient condition for the perception at a corresponding level of accuracy of the shapes composing the surface (see Flock, 1964c).

Secondly, *S*s first made slant judgments, then made shape judgments. The shape judgments, following the slant judgments, possibly benefited from the prior exposure to the display. This procedure might have introduced a bias in favor of making shape judgments that were more accurate than slant judgments, just as by reason of a bias derived from a different asymmetry Stavrianos (1945) found better shape than slant judgments. In fact, however, slant judgments were accurate when conditions warranted it, but shape judgments under all conditions were inaccurate, varying with the retinal shapes, showing no shape constancy, and failing to covary with the perceived slants.

GENERAL DISCUSSION

Regardless of the fact that velocity was altered by a factor of 4:1 in Exp. I–III; regardless of the substitution of one electrostatic texture for a second as in Exp. I and IV; regardless of the use of irregular and unfamiliar textures in Exp. I–IV and regular, familiar textures in Exp. V–VII; whether *S*s have practice or not with the displays as in Exp. I and VIII; so long as motion is discernibly present (Exp. VIII vs. IX) and field of view is at least 17° or 35° (Exp. X), the relationship between observed and physical slant found in Exp. I can be successfully replicated under a wide variety of conditions. Only in the case of the low-density, ½-in. regular surface did there occur a failure to replicate the findings from the reference experiment, Exp. I. Moreover, *S*s' success in judging the slants of the irregular, unfamiliar surfaces (Exp. I–IV and VIII) did not seem to depend on their ability to be able to judge the shape of at least one texture element on the surface (Exp. XI) or on the presumptive familiarity of the surface's texture elements (Exp. I–IV vs. V–VII).

These findings seem to support the general hypothesis of this paper: that accurate slant judgments may depend on a pick-up from the light to the eye of relational variables which in turn specify correctly the slant of distal surfaces (see Flock, 1962, for a specification of these theoretical variables). That hypothesis omits any reference to the particular stuff of which a surface is made. Any moving surface texture at a particular slant, regardless of its distance or velocity, will project exactly the same optical variable specifying its slant as any other surface. Change the slant of the surface and only the value given to the variable will be changed. In the experiments above the textural stuff composing some of the surfaces was presumably novel to *S*s. But on the basis of the empirical findings it would seem that the higher-order optical variables specifying the surface's slant were not novel.

Hochberg and McAlister (1955) have suggested that there are at least two strategies in attempting to explain the problems of monocular space perception. One, the traditional strategy, is to postulate subjectivist variables. The other, whose major proponent in perception today seems to be Gibson, is to attempt to identify the variables of stimulation that on the one hand informatively specify the spatial characteristics of distal objects and on the other induce correct perceptions of the objects. If we are to advance our knowledge about these optical variables, presumably it will be increasingly necessary that we attend to the relationships between the characteristics of physical objects and surfaces on the one hand and the variables in the patterned light corresponding to those physical characteristics on the other—what Gibson (1961) has recently referred to as ecological optics. For an example of this kind of analysis and the empirical predictions it yields, see Flock (1964b).

REFERENCES

BECK, J., & GIBSON, J. J. The relation of apparent shape to apparent slant in the perception of objects. *J. exp. Psychol.*, 1955, **50,** 125–133.

BRUNSWIK, E. Distal focussing of perception: Size-constancy in a representative sample of situations. *Psychol. Monogr.*, 1944, **56** (1, Whole No. 254).

BRUNSWIK, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

EPSTEIN, W., & PARK, J. N. Shape constancy: Functional relationships and theoretical formulations. *Psychol. Bull.*, 1963, **60,** 265–288.

FLOCK, H. R. *The monocular perception of surface slant.* (Doctoral dissertation, Cornell University), Ann Arbor, Mich.: University Microfilms, 1962, No. 62-2514.

FLOCK, H. R. Irregular textures for the point-source shadow-projector. *Amer. J. Psychol.*, 1964, in press. (a)

FLOCK, H. R. A possible optical basis for monocular slant perception. *Psychol. Rev.*, 1964, **71,** in press. (b)

FLOCK, H. R. Three theoretical views of slant perception. *Psychol. Bull.*, 1964, **62,** in press. (c)

GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.

GIBSON, J. J. Ecological optics. *Vis. Res.*, 1961, **1,** 253–262.

GIBSON, J. J., & GIBSON, E. J. Perceptual learning: Differentiation or enrichment? *Psychol. Rev.*, 1955, **62,** 32–41.

GIBSON, J. J., & GIBSON, E. J. Continuous perspective transformations and the perception of rigid motion. *J. exp. Psychol.*, 1957, **54,** 129–138.

GIBSON, J. J., OLUM, P., & ROSENBLATT, F. Parallax and perspective during aircraft landings. *Amer. J. Psychol.*, 1955, **68,** 372–385.

GRANT, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53,** 141–154.

HELMHOLTZ, H. *Physiological optics.* Vol. 3. (3rd ed. trans. by J. P. Southall) New York: Optical Society of America, American Institute of Physics, 1925.

HOCHBERG, J. E., & MCALISTER, E. Relative size vs. familiar size in the perception of represented depth. *Amer. J. Psychol.*, 1955, **68,** 294–296.

KOFFKA, K. *Principles of Gestalt psychology.* New York: Harcourt, Brace, 1935.

PURDY, W. C. *The hypothesis of psychophysical correspondence in space perception.* (Doctoral dissertation, Cornell University), Ann Arbor, Mich.: University Microfilms, 1958, No. 58-5594.

RYAN, T. A. Multiple comparisons in psychological research. *Psychol. Bull.*, 1959, **56,** 26–47.

RYAN, T. A. Significance tests for multiple comparison of proportions, variances, and other statistics. *Psychol. Bull.*, 1960, **57,** 318–328.

SNEDECOR, G. W. *Statistical methods.* (4th ed.) Ames, Iowa: Iowa State Coll. Press, 1948.

STAVRIANOS, B. K. The relation of shape perception to explicit judgments of inclination. *Arch. Psychol., N. Y.*, 1945, No. 296.

(Received July 18, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 573–576

HANDLING-GENTLING AS A POSITIVE SECONDARY REINFORCER[1]

SALLY E. SPERLING AND FRED P. VALLE

University of California, Riverside

Food-deprived rats were given pretraining experience in which handling was paired with eating. During 12 acquisition sessions they learned to approach the compartment of a 2-compartment box from which they were removed and stroked, demonstrating handling to be an effective positive secondary reinforcer. 2 other groups, one given pretraining handling without deprivation and the other neither handled nor deprived, showed no systematic change in entrances into their critical compartments. The 2 latter groups differed from each other in general activity level, with the prehandled Ss being more active.

Recent experiments by Candland and associates (Candland, Faulds, Thomas, & Candland, 1960; Candland, Horowitz, & Culbertson, 1962) have cast doubt upon the efficacy of pre-experimental taming procedures in reducing or neutralizing the rat's seeming initial aversion to handling, as indexed by struggling, biting, urinating, defecating, etc., and have raised the question of the extent to which rat learning data might be contaminated by the negatively reinforcing properties of handling. These investigators have shown (Candland et al., 1960) that both pre-experimentally gentled and nongentled rats learned to avoid an end box from which they were removed and stroked, choosing in preference a box where they were detained, and that they reversed this discrimination when the critical compartment was reversed. In a second study (Candland et al., 1962), rats learned to avoid the black compartment of a two-compartment shuttle box from which they were removed and stroked for 20 sec. and showed retention of this response for at least 15 days.

One of the primary differences between the usual pre-experimental procedures and those used by Candland et al., is that pre-experimental handling is usually accompanied by the establishment of a primary deprivation schedule while no such deprivation manipulations were utilized in the above experiments. The present experiment, therefore, was designed to examine learning with handling as a reinforcer following a pretraining period in which handling was paired with eating for food-deprived rats in the expectation that this pretraining, which is more like the usual pre-experimental experience, would more effectively neutralize the aversive properties of handling than handling alone.

METHOD

The *S*s were 30 male and female hooded rats bred and maintained in the laboratory colony. They had been weaned at 21 days of age, maintained in group cages of from 4 to 6 rats until sexing at age 35 days, and maintained in like-sized, same-sexed groups until age 48 days. All prior handling of *S*s had been done by lifting them by the upper portion of the tail.

At age 48 days all *S*s were removed from group cages, by lifting them by the tail, and housed individually in the same rack. One group of six males and four females was

---

[1] This study was supported by a grant, G-24168, from the National Science Foundation to the first author.

maintained on ad-lib. food and water but not handled for an ensuing 10-day period (Group NH). A second group of six males and four females was also maintained on ad-lib. food and water but each *S* was picked up from its cage once daily for the first 7 days of pretraining and at 12-hr. intervals for the last 3 days and stroked for 5 min. following the procedure described by Weininger (1956); *E* held and stroked *S* from head to tail approximately 50 times per minute (Group H). The third group of seven males and three females had continuous access to water but were fed dry lab chow mixed with water from a small glass cup while *E* held and stroked them following the above procedure. These *S*s (Group HD) were fed once a day for the first 7 days (8 gm. per day) and at 12-hr. intervals (4 gm. per feeding) for the last 3 days.

Twelve hours after the last pre-experimental handling for Groups H and HD, each *S* in each of the three groups was placed in the apparatus for a 5-min. adaptation period, which was divided for scoring purposes into 15 20-sec. intervals for a measure of pre-experimental activity.

Pre-experimental handling for Groups H and HD was discontinued at this time; Group HD *S*s were maintained on a 12-hr. deprivation schedule, receiving 4 gm. of dry food mixed with water after each acquisition session. For the 12 acquisition sessions which followed for each *S* at 12-hr. intervals, initial placement into the apparatus was always into the same compartment—*S*'s critical compartment determined semirandomly before adaptation. If *S* was in this compartment at the end of any of 15 20-sec. intervals, he was removed by *E*, held in a gloved hand, stroked with an ungloved hand for 20 sec. at the rate of approximately 50 strokes per minute in a head-to-tail direction, and replaced in the apparatus depending on a random schedule for posthandling placements.

The apparatus was a plywood rectangular box, 22 in. long, 7 in. wide, and $8\frac{3}{4}$ in. high, divided into two compartments, one painted flat black and the other flat white, by a plywood partition with a 4 in. high × 3 in. wide top-swinging wire mesh door. The floor of the apparatus was of $\frac{1}{8}$ in. diameter metal bars placed $\frac{1}{2}$ in. apart; the top was hinged wire mesh. Overhead illumination was 29.5 ft-c; illumination in the black and white compartments was 23.4 ft-c and 24.5 ft-c, respectively.

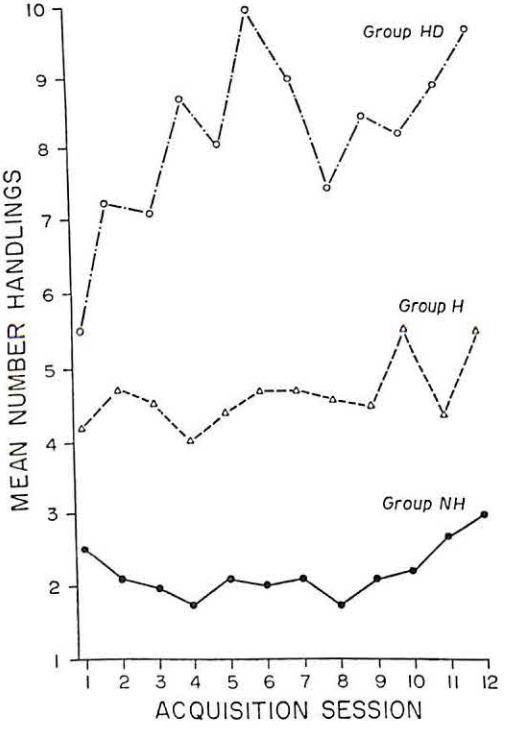


FIG. 1. Mean number of handlings per session of 15 20-sec. intervals by groups.

Results[2]

During the adaptation period, Groups NH, H, and HD were observed to have changed compartments at the end of successive intervals an average of 1.5, 2.1, and 2.4 times, respectively. An analysis of variance yielded an F value less than unity, indicating that the groups could not be considered to be different during adaptation.

The mean number of handlings per group per session is depicted in Fig. 1. Analysis of variance indicated that there were differences among groups, $F\ (2, 27) = 13.92$, $p < .001$, and sessions, $F\ (11, 297) = 2.13$, $p < .025$. Tukey 95% confidence intervals were computed for the overall group means.

[2] Neither the analyses by sex differences nor critical compartment differences approached statistical significance. Therefore, these data have been combined for the respective groups.

By this analysis, all three group means differed significantly from each other.

Tukey 95% confidence intervals were also computed for the group means for Session 1 and Session 12, using the respective mean squares for within treatments. None of the group means for Session 1 could be considered to be different; for Session 12, Group HD received significantly more handlings than either Groups NH or H which did not differ from each other.

A within-group comparison was made between Sessions 1 and 12 by Tukey 95% confidence intervals. This analysis showed that the mean number of handlings could not be considered different for Group NH or Group H, but that Group HD was handled significantly more often during Session 12 than during Session 1.

DISCUSSION

The data for Group HD showed that these *S*s learned to approach the compartment from which they were removed and handled. Since no primary reinforcement was given during acquisition sessions, it must be concluded that handling was an effective positive secondary reinforcer for these *S*s, and, since Group H *S*s did not exhibit this effect, that it is the combination of deprivation and handling that accounts for the Group HD data. Whether the preacquisition pairing of feeding and handling is a necessary condition for the establishment of handling as a secondary reinforcer cannot be determined from these data, but it is clear that it is at least a sufficient condition and that the secondary reinforcer is strong and persistent under this procedure.

With respect to Groups H and NH, the present data do not replicate the findings of Candland et al. (1960, 1962). There was no evidence of learning based on the aversive properties of handling or, for that matter, any learning at all, and the pre-experimentally handled only group was different from the not handled group in terms of total number of handlings during acquisition.

The difference between Groups NH and H seems best accounted for by the hypothesis that the effect of handling on Group NH *S*s was to decrease or depress their general activity level in comparison to Group H, and, of course, Group HD. The data showed that the groups were exhibiting approximately equal, and low, amounts of activity during the adaptation period when they were not handled. Leaving out Group HD whose acquisition data confound spontaneous activity and learning, comparisons between Groups NH and H on the number of spontaneous crossings during the first acquisition session, i.e., ignoring intervals where *E* had replaced *S* in a different compartment from which he was removed, showed that Group H *S*s exhibited almost three times as many crossings as Group NH which showed almost no change from the adaptation period. By the end of acquisition, Group NH's spontaneous activity level had increased slightly but was still less than half of Group H's. The Group NH *S*s, then, seemed more likely to stay where they were, whether in their critical compartment or not, than the Group H *S*s who moved about the apparatus more freely.

This conclusion is supported by an experiment by Weininger (1956) in which it was found that pre-experimental gentling of albino rats resulted in more ambulatory movement in an open field than was observed in *S*s not pre-experimentally handled. The data of Barry (1957) and Bernstein (1957) are also consistent with this interpretation.

It seems most likely that the usual pre-experimental procedures where taming and feeding are separated in time are effective in neutralizing the initially aversive effects of handling and that these *S*s, like Group H *S*s, do not react to handling during the experiment proper in any systematic fashion.

REFERENCES

BARRY, H., III. Habituation to handling as a factor in retention of maze performance in rats. *J. comp. physiol. Psychol.*, 1957, **50,** 366–367.

BERNSTEIN, L. Effects of variation in handling upon learning and retention. *J. comp. physiol. Psychol.*, 1957, **50,** 162–167.

CANDLAND, D. K., FAULDS, B., THOMAS, D. B., & CANDLAND, M. H. The reinforcing value of gentling. *J. comp. physiol. Psychol.*, 1960, **53,** 55–58.

CANDLAND, D. K., HOROWITZ, S. H., & CULBERTSON, J. L. Acquisition and retention of acquired avoidance with gentling as reinforcement. *J. comp. physiol. Psychol.*, 1962, **55,** 1062–1064.

WEININGER, O. The effects of early experience on behavior and growth characteristics. *J. comp. physiol. Psychol.*, 1956, **49,** 1–9.

(Received July 15, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 577–580

EFFECT OF SPATIAL SEPARATION OF STIMULUS, RESPONSE, AND REINFORCEMENT ON SELECTIVE LEARNING IN CHILDREN[1]

WENDELL E. JEFFREY AND LESLIE B. COHEN

University of California, Los Angeles

60 children of 2 age levels were randomly assigned for training under 5 possible combinations of spatial contiguity or separation of stimulus, response, and reinforcement. 2 levels of stimulus difficulty were used. Children learned only in the 2 conditions where stimulus and response were contiguous. Previous evidence indicating that monkeys learned under stimulus-reward contiguity was shown probably to be a matter of transfer of training from previous training under conditions of stimulus-response contiguity in a repeated-measures design. Age was not an important variable in learning the easy discrimination, and the difficult discrimination proved to be too difficult for both age groups.

A number of studies of selective learning have shown that spatial separation between stimulus and response markedly retards learning in primates as well as children (Meyer, Polidora, & McConnell, 1961; Murphy & Miller, 1955, 1959; Stollnitz & Schrier, 1962).

Murphy and Miller (1958) also manipulated the locus of reward and found that monkeys could learn with the stimulus and response separated as long as the reward was contiguous with the stimulus. However, inasmuch as each monkey served in all conditions of stimulus, response, and reward separation, their results may have reflected transfer effects from previous training.

The present experiment was designed to re-evaluate the problem of spatial contiguity of stimulus, response, and reward. Nursery school children were used in independent groups, thereby eliminating any potential facilitation through prior training. Two age groups and two levels of stimulus similarity were also used. In addition, after the initial experiment was completed, the effect of previous training on transfer to the condition where stimulus and reward are contiguous and response is separated was evaluated.

METHOD

Subjects.—Sixty nursery school children served as *S*s. They were equally divided into two age categories: a Young group with mean CA of 41 mo. and a range of 36–47 mo., and an Old group with mean CA of 53 mo. and a range of 48–58 mo. Adequate rapport was established with each child before he participated in the experiment.

Apparatus.—A trailer parked adjacent to the play yard of the nursery school housed a modified Wisconsin General Test Apparatus (WGTA). The test tray, painted medium gray, held five wooden cubes, 2½ in. on each side, located as indicated schematically in Fig. 1. Blocks in the upper left (UL), center (C), and upper right (UR) of the figure were also painted medium gray while stimulus blocks, LL and LR, were either white and black, or light and dark gray. A felt-lined reward cup, 1½ in. in diameter and ⅝ in. deep, hidden under each block, could hold token reinforcers which were exchanged for a bag of trinkets at the conclusion of the experiment. The tokens were collected in a small green plastic hat. Cups UL and UR were 11 in.

---

[1] This research was supported in part by grants from the National Science Foundation (G 10041) and the University of California Faculty Research Committee.

We are extremely grateful to Ruth Brody of the Children's World Nursery School who gave her fullest cooperation and wise counsel.

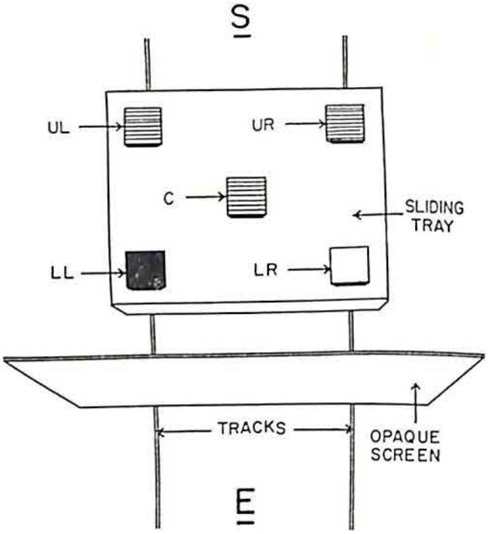


FIG. 1. Schematic diagram of layout of WGTA.

apart as were Cups LL and LR. There was an 8-in. separation between UL and LL, and UR and LR. Cup C was also 8 in. removed and equidistant from each of the others. A gray cloth screen which could be opened and closed by *E* prevented *S* from observing the baiting of the tray.

Procedure.—The *S*s in the two age groups were randomly assigned to one of five groups differing in spatial separation of stimulus (S), response (R), and reward (Re). The five spatial-separation conditions were as follows. Cond. I: contiguous S, R, Re; Cond. II: S, R contiguous, Re separated; Cond. III: S, Re contiguous, R separated; Cond. IV: R, Re contiguous, S separated; Cond. V: S, R, Re separated. One half of each group (three *S*s) was trained on the easy, black-white discrimination; the other half was trained on the more difficult dark gray-light gray discrimination. Training proceeded for 60 trials. The rewarded stimulus was counterbalanced by having the lighter of the two stimuli correct for one half the *S*s, the darker stimulus being correct for the remaining half. The positions of the Lighter and Darker stimuli were systematically interchanged throughout the trials according to the sequence, LDDLLLDDDLLD, which indicated the stimulus to be placed to *E*'s left.

The *S* was seated in front of the WGTA and told that if he filled the green hat in front of him with the little black beads he could win all the trinkets. Following this *E* sat behind the WGTA out of *S*'s view. After hiding the bead under the appropriate block for the experimental condition, the tray was pushed forward. The *S*'s attention was called to the two response blocks and he was also told not to touch any of the remaining blocks. The *S* was then given a sample trial in which both a reinforced and nonreinforced response were required.

In Cond. I removal of the correct stimulus block (LL or LR) revealed the reward. In Cond. II *S*'s response was again to the stimulus block but reinforcement was delivered by *E*'s removal of Block C if *S*'s choice was correct. Otherwise the tray was withdrawn. In Cond. III *S* responded to UL or UR and *E* removed the stimulus block (LL or LR) on the same side as *S*'s response, uncovering a bead if *S* had responded to the correct side. In Cond. IV *S* found the reinforcer under the response block (UL or UR) on the same side as the correct stimulus block. In Cond. V *S* responded to UL or UR and was reinforced at C as in Cond. II.

Once *S* had picked up one block and returned it, the tray was pulled back by *E*, the screen closed, and the blocks put in position for the next trial. The *E* then reopened the screen to return the tray. Each 60-trial testing session averaged about 20 min. At the conclusion of 60 trials every *S* was told that he now had a sufficient number of beads, and was given the bag of trinkets.

RESULTS AND DISCUSSION

The black-white stimulus pair led to a higher frequency of correct responses than the dark gray-light gray pair, $F\,(1, 48) = 17.89$, $p < .001$. The older *S*'s performance did not differ reliably from that of the younger *S*s and no interaction reached statistical significance.

Frequency curves of acquisition for the five treatment conditions, ignoring both stimulus similarity and age, are shown in Fig. 2. The difference between the five conditions on number of correct responses was statistically significant, $F\,(4, 40) = 8.97$, $p < .001$. Duncan's multiple-range test indicated that Cond. I and II were reliably different ($p < .001$) from III, IV, and V, but were not significantly different from each other. The *S*s under Cond.

III, IV, and V, did not significantly differ from one another.

Murphy and Miller (1958) concluded that stimulus-reward contiguity significantly facilitated learning, yet the children under this condition in the present experiment did not learn. In order to determine whether Murphy and Miller's results might be attributable to transfer of training, three children were randomly selected from the black-white cue groups of Cond. I, II, IV, and V. Each *S* was first given 24 retraining trials under his original training condition followed by the appropriate instructions and 48 training trials on Cond. III. Those *S*s who had had prior training under Cond. I and II, transferred quite easily to Cond. III obtaining a mean score of 89.6% correct over 48 trials, whereas those trained previously under Cond. IV and V averaged only 60.6% correct. The difference between these groups was significant, $F(1, 10) = 6.19$, $p < .05$.

There is some evidence that the acquisition of appropriate or inappropriate observing responses may be at least partially responsible for the varying degrees of difficulty of the five tasks. Schuck (1960) has shown that monkeys will tend to restrict their observations to the region of the manipulanda, thus precluding a view of the stimuli when spatially separated from the response. Indeed, if a panel which prevents observation of the manual response (Otteson, Sheridan, & Meyer, 1962) is inserted in the apparatus, learning to discriminate remote stimuli is facilitated, presumably because the likelihood that the stimuli will be observed is increased. Thus, when the stimulus and response are spatially separated (Cond. IV or V), *S* tends to restrict his view to the area of the response. In Cond. III, however, observing responses should be highly likely to the stimuli contiguous with the reward, but inasmuch as that observation follows the response it apparently does not affect performance unless *S* has previously had experience with the stimuli and responses spatially contiguous. We must conclude, then, that with experimentally naive children, stimulus-response contiguity is a necessary condition for learning. Whether they can learn with the stimulus and response separated when the manipulanda are not observable has yet to be determined.

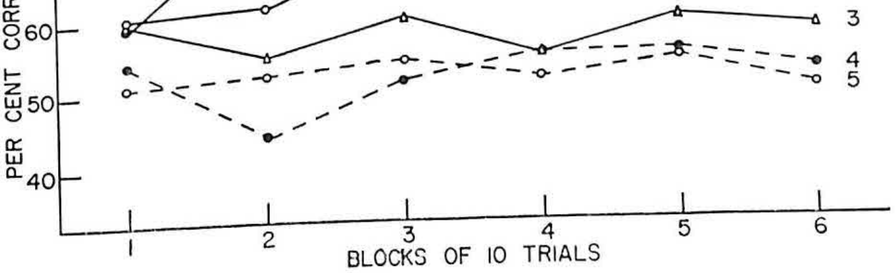


FIG. 2. Percentage of correct responses under the various conditions: (1) cue, response, reward; (2) cue, response . . . reward; (3) cue, reward . . . response; (4) cue . . . response, reward; (5) cue . . . response . . . reward.

REFERENCES

MEYER, D. R., POLIDORA, V. J., & McCONNELL, D. G. Effects of spatial S-R contiguity and response delay upon discrimination performance by monkeys. *J. comp. physiol. Psychol.*, 1961, **54**, 175–177.

MURPHY, J. V., & MILLER, R. E. The effect of spatial contiguity of cue and reward in the object-quality learning of rhesus monkeys. *J. comp. physiol. Psychol.*, 1955, **48**, 221–229.

MURPHY, J. V., & MILLER, R. E. The effect of the spatial relationship between the cue, reward, and response in simple discrimination learning. *J. exp. Psychol.*, 1958, **56**, 26–31.

MURPHY, J. V., & MILLER, R. E. Spatial contiguity of cue, reward, and response in discrimination learning by children. *J. exp. Psychol.*, 1959, **58**, 485–489.

OTTESON, M. I., SHERIDAN, C. L., & MEYER, D. R. Effects of stimulus-response isolation on primate pattern discrimination learning. *J. comp. physiol. Psychol.*, 1962, **55**, 935–938.

SCHUCK, J. R. Pattern discrimination and visual sampling by the monkey. *J. comp. physiol. Psychol.*, 1960, **53**, 251–255.

STOLLNITZ, F., & SCHRIER, A. M. Discrimination learning by monkeys with spatial separation of cue and response. *J. comp. physiol. Psychol.*, 1962, **55**, 876–881.

(Received July 20, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 581–586

PARTIAL REINFORCEMENT EFFECT AND EXTINCTION AS A FUNCTION OF FRUSTRATION AND INTERFERING RESPONSES[1]

LANGDON E. LONGSTRETH

University of Southern California

4 groups of rats were trained and extinguished in a straight alley. 3 groups were 100% reinforced. 1 of these was intermittently fed in the start box during training (SB+). Another served as a control for handling and confinement procedures, but was not fed in SB (SB−). The 3rd group was given conventional 100% reinforcement training (100-C). A 4th group was given conventional 50% reinforcement training (50-C). In general, the results indicated: (a) the typical PRE, with 50-C most resistant and 100-C least resistant to extinction, (b) SB+ more resistant to extinction than SB− or 100-C, (c) no relationship between goal-box behavior and resistance to extinction, and (d) faster extinction for Ss who avoided the goal box during extinction. These results were interpreted as supporting frustration theory and as incompatible with contiguity-interference theory.

Two prominent explanations of the partial reinforcement effect (PRE) are Amsel's frustration theory (Amsel, 1958) and Weinstock's interference-habituation theory (Weinstock, 1954, 1958). According to Amsel's position, extinction will be retarded to the extent that the instrumental response (R_{in}) is conditioned during acquisition to frustration-produced stimuli (s_f) which are present during extinction. The usual way of providing such cues is to partially reinforce R_{in}, thus presumably frustrating *S* during acquisition. Another way would be to intermittently feed *S* in the start box (SB), while reinforcing him in the goal box (GB) 100% of the time. Assuming that conditioned frustration responses would then be elicited by placement in an empty SB, and allowing R_{in} to occur contiguously with such placement, conditions would be created which should provide the necessary s_f-R_{in} conditioning. These conditions were provided in the present study, with the prediction that such *S*s would be more resistant to extinction than control *S*s not fed in SB but 100% reinforced in GB.

According to Weinstock's position, extinction is produced by responses in the empty GB which generalize to the SB or alley and compete with R_{in}, producing extinction. Thus a rat which backs or turns around in GB should extinguish faster than a rat which does not engage in these or other incompatible responses. These responses are "habituated," however, by nonreinforced acquisition trials. Thus partially reinforced *S*s emit such responses at a low rate during extinction, resulting in greater resistance to extinction.

The variety of goal-box behavior makes it difficult to record incompatible responses, thus perhaps accounting for the infrequency with which such data are reported. In the present study, the apparatus was constructed in such a way as to eliminate most gross behavior in GB other than

---

[1] This research was supported by grants from the National Institutes of Health (M-6331) and the National Science Foundation (G-16301). The author is indebted to Norman Gold, who ran *S*s in this experiment.

backing up or turning around. This was accomplished by restricting the inside dimensions of GB to 2 in. in width and height. According to Weinstock's theory, it was therefore predicted that resistance to extinction would be inversely related to amount of backing and turning-around behavior. It was also predicted that these responses would occur with decreasing frequency over the nonreinforced acquisition trials of a partially reinforced group, and that these *S*s would enter extinction with a lower rate of occurrence for such responses than 100% reinforced *S*s.

METHOD

Subjects

Sixty naive albino rats of the Wistar strain who had been bred in the departmental colony and were about 100 days old at the beginning of the study served as *S*s. They were randomly assigned to four groups, with 13, 14, 16, and 17 *S*s in each group. The data of another 11 *S*s were discarded because of failure to run or eat during preliminary trials. In order to increase precision, the data of 12 female *S*s who exhibited considerable turning-around behavior in GB were discarded except when specifically mentioned, leaving final *N*s of 12, 12, 14, and 10.

Apparatus

The apparatus consisted of a gray SB, alley, and GB. Height and width were 2 in. throughout, and length of the three components was 10 in., 4 ft., and 18 in., respectively. Two doors separated the three components, and hinged false floors 3 in. from the ends of the alley activated a circuit leading to a Standard Electric clock, providing an automatic measure of running time. A second clock was activated manually, approximately simultaneously with opening of the SB door, and terminated automatically with depression of the first false floor, thus providing a measure of starting time. The cover of the entire apparatus was clear plastic, making it possible to observe *S* in all parts of the apparatus.

Procedure

Habituation.—All *S*s were placed on a 23-hr. food deprivation schedule 5 days before the experiment began. Water was available ad libitum. This schedule was maintained throughout the experiment, and *S*s were about 22 hr. hungry on all acquisition and extinction trials. Food consisted of ground wet Purina chow.

All *S*s received five preliminary trials, with food scattered throughout the alley and GB on the first three trials and restricted to a small metal pan at the end of GB on the last two trials.

Acquisition.—Thirty acquisition trials were administered to all *S*s, with two or three trials per day and an intertrial interval ranging from 1 to 2 hr. Confinement in GB was for 30 sec.

Three groups were 100% reinforced in GB. One of these groups was fed in SB on a randomly determined 15 of the 30 trials. The procedure on these trials was as follows: *S* was fed 15 sec. in SB, usually consuming all the available food in that time, and was then removed to a waiting cage while SB was cleaned of food particles. Replacement followed immediately, and the SB door was opened 5 sec. later, allowing access to the alley. On all other trials, *S*s of this group (SB+), as well as *S*s of all other groups, were confined in SB for 5 sec.

A second 100% reinforcement group was treated exactly the same except that SB was always empty (SB−). A third 100% reinforcement group was given conventional treatment (100-C). That is, SB was always empty and confinement was always 5 sec. A fourth group was given conventional 50% reinforcement training. The schedule of GB reinforcements was the same as the schedule of SB feedings for Group SB+. Thus these two groups were exposed to the same intermittent reinforcement schedule, one group in SB and the other in GB.

Extinction.—All groups received the same treatment during extinction: 30 nonreinforced runs from SB to GB. All *S*s were confined in SB for 5 sec. and in GB for 30 sec. The door to GB was closed after *S* entered it to prevent retracing. If *S* did not traverse the alley in 30 sec., he was removed and a running time of 30 sec. recorded for that trial. Three or four trials were given per day, with an intertrial interval ranging from 1 to 2 hr.

Response measures.—Five response measures were recorded: starting time, running time, frequency of turning around in GB, amount of backing in GB, and avoidance of GB. Amount of backing was rated on a 4-point scale: a score of 3 represented 12 or more in. of backing, a score of 2 represented between 2 and 12 in. of backing, a score of 1 represented between 0 and 2 in. of backing, and if *S* did not move at all, even by pulling his head back from the food pan, a score of 0 was recorded. Independent ratings by two *O*s in a preliminary study indicated satisfactory reliability, with only 2 of 28 pairs of ratings showing disagreement.

Avoidance of GB was recorded if *S* traversed the alley within 30 sec., but did not enter GB. Turning around (TA) was recorded if *S* turned around after entering GB.

RESULTS AND DISCUSSION

Resistance to extinction.—Means of median starting and running times for each block of five extinction trials are presented in Fig. 1 and 2. Acquisition differences for the four groups were negligible, as is suggested by means on the first block of extinction trials. Both figures indicate similar trends: (*a*) the typical PRE, with Group 100-C extinguishing fastest and 50-C slowest, and (*b*) SB+ more resistant to extinction than SB− or 100-C, particularly with starting times.

Analysis of variance of starting times yielded a significant Groups × Trials interaction, *F* (15, 220) = 3.90, $p < .001$, suggesting the four groups did not extinguish at the same rate. Heterogeneous within-cell variances, due mainly to less variability for Cond. 50-C, suggested the need for another type of evaluation. A median test was performed over all 30 extinction trials, dividing each group into those above and below the general median. The results were χ^2 (3) = 12.59, $p < .01$, confirming the analysis of variance results.

In order to determine if there was systematic extinction variance not accounted for by the PRE, a second analysis of variance was performed, excluding the data of Group 50-C. A significant Groups × Trials interaction was obtained, *F* (10, 175) = 4.39, $p < .001$. A final analysis of Groups SB+ and SB− alone, employing the error term from the previous three-group analysis, also yielded a signifi-

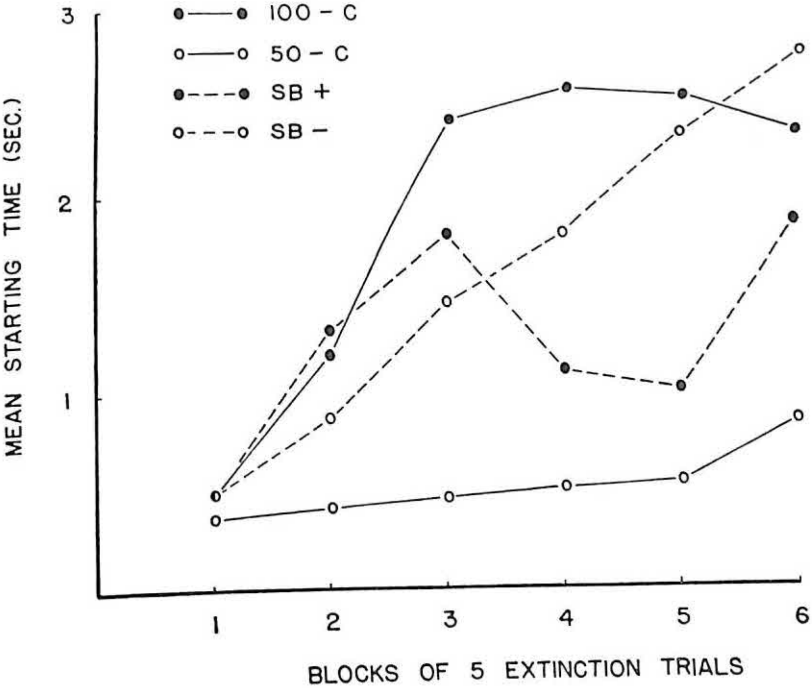


FIG. 1. Mean starting time in extinction as a function of experimental conditions.

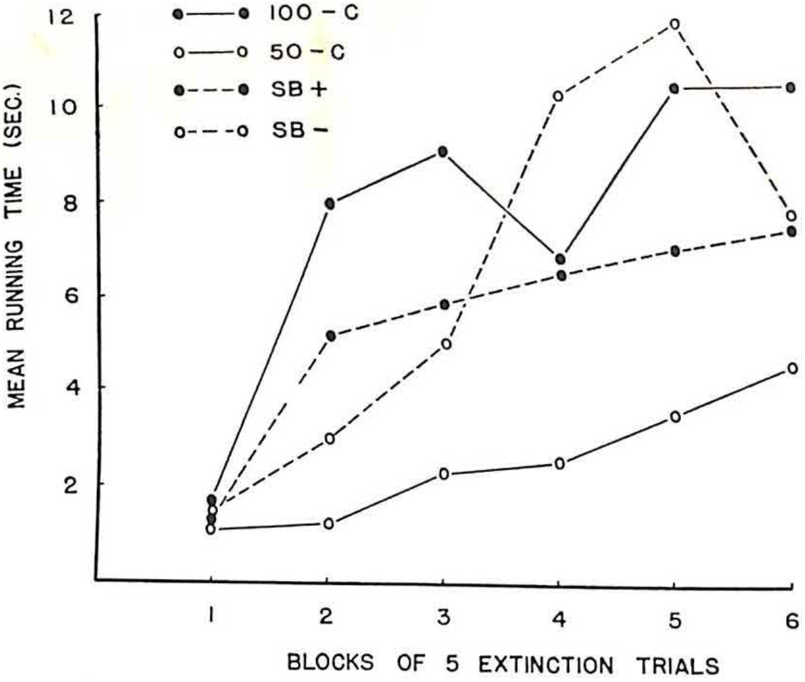


FIG. 2. Mean running time in extinction as a function of experimental conditions.

cant Groups × Trials interaction, $F(5, 175) = 3.66$, $p < .005$. It is therefore concluded that intermittent SB feedings during training resulted in greater resistance to extinction than the absence of such feedings.

Similar analyses of running times indicated that Groups SB+, SB−,

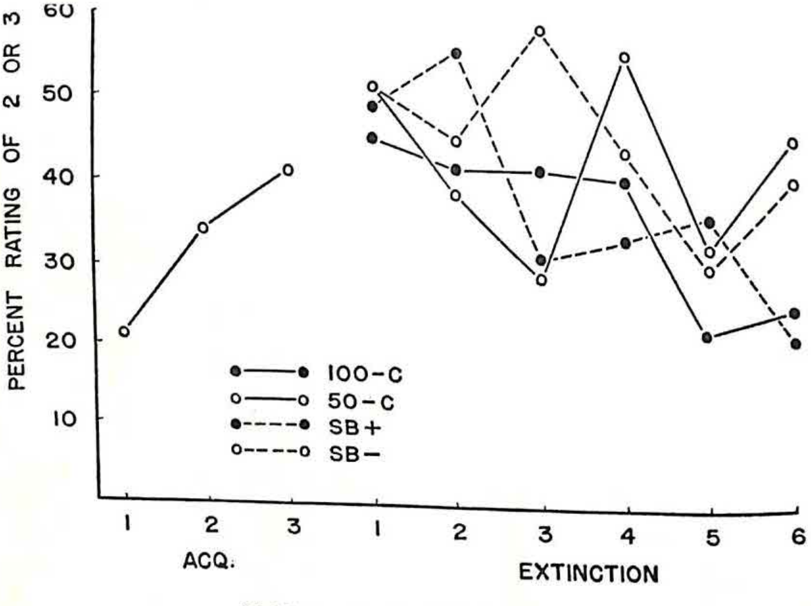


FIG. 3. Mean proportions of GB backing ratings of 2 and 3 as a function of experimental conditions.

and 100-C all extinguished at the same rate. Within Amsel's theory, this discrepancy is not entirely unexpected. It need only be observed that the *beginning* of R_{in} was more contiguous with empty SB confinement than later segments of R_{in}. Thus one would expect stronger conditioning of s_f to the starting response than to the running response.

Goal-box behavior.—Backing behavior during the 15 nonreinforced acquisition trials (Group 50-C) and 30 extinction trials is presented in Fig. 3, in blocks of five trials. This figure indicates the mean proportion of trials on which backing behavior in excess of 2 in. (a rating of 2 or 3) occurred. Only trials on which *S*s entered GB were used in calculating these means. Figure 3 indicates that backing behavior increased (nonsignificantly) over the 15 nonreinforced acquisition trials for Group 50-C, and that amount of backing behavior during early extinction trials was about the same for all four groups. These data appear to be inconsistent with Weinstock's explanation of the PRE. Turning-around proportions were not calculated for the four groups because only two or three *S*s per group manifested such behavior.

The relationship between backing behavior and resistance to extinction was evaluated by first matching *S*s within groups in terms of sex and number of trials GB was entered. Three matched pairs were then picked from each group who differed most in the proportion of trials on which a backing rating of 2 or 3 was recorded. The two resulting groups of 12 *S*s were thus closely matched on sex (2 females, 10 males) and number of GB entrances (234 and 235), and differed only in amount of backing behavior (median proportions of ratings of 2 or 3 equaled .14 and .58). Starting and running times both failed to yield a difference between these two groups that even approached significance.

The relationship between TA and resistance to extinction was similarly evaluated, using the data of females, since only females manifested this behavior. Two groups of 10 females were formed, matched on number of trials GB was entered and differing on proportion of trials that TA was recorded (mdn. proportions = .14 and .46). As with backing behavior, neither starting nor running times indicated a difference between these two groups that approached significance.

The relationship between GB avoidance and resistance to extinction was evaluated in the same general way. Two groups of 12 *S*s, matched on sex and number of trials on which the alley was traversed, differed on the proportion of trials on which GB was avoided (mdn. proportions = .21 and .45). Running times were significantly different for these two groups, with faster extinction for Group .45, $F\,(1, 22) = 6.33$, $p < .05$. Starting times, however, revealed no significant difference. It was also found that Group 50-C engaged in significantly less GB avoidance than the other three groups, $F\,(3, 44) = 4.23$, $p < .05$.

The lack of a relationship between backing and TA behavior and resistance to extinction is not only inconsistent with contiguity-interference theory, but also appears to be at odds with data reported by Adelman and Maatsch (1955). These investigators found that turning around in GB and retracing the alley resulted in faster extinction than GB confinement. However, it must be recalled that turning around resulted in escape from GB, whereas in the present study escape was impossible. Within a frustration point of view, turning around followed by escape from GB would be reinforced by a reduction in frustration-produced drive, and therefore would be learned, resulting in faster extinction. With no possibility of escape, frustration reduction would not occur, and thus no learning would take place. Thus a *reinforcement-interference* theory, such as Amsel's, would fit both sets of data. Such a point of view is further suggested by the present finding of a positive relationship between GB avoidance and extinction rates. Avoidance behavior, unlike backing and TA, resulted in avoidance of GB cues, and, presumably, in reduction of conditioned frustration responses. Wagner (1963) has recently presented data more directly supporting the motivational properties of these hypothetical responses. He found that cessation of cues previously associated with nonreward served as reinforcement for a hurdle-crossing response.

REFERENCES

ADELMAN, H. M., & MAATSCH, J. L. Resistance to extinction as a function of the type of response elicited by frustration. *J. exp. Psychol.*, 1955, **50**, 61–65.

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

WAGNER, A. R. Conditioned frustration as a learned drive. *J. exp. Psychol.*, 1963, **66**, 142–148.

WEINSTOCK, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **47**, 318–322.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hour intertrial interval. *J. exp. Psychol.*, 1958, **56**, 151–158.

(Received July 26, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 587–590

DEMONSTRATION OF ACQUIRED DISTINCTIVENESS OF CUES USING A PAIRED-ASSOCIATE LEARNING TASK[1]

ERWIN M. SEGAL[2]

University of Minnesota

152 students performed 2 tasks. First, they learned to identify stimulus configurations using 3, 9, or 18 different responses. The response referred to values of 1, 2, or 3 stimulus dimensions. There were significant differences between groups responding to different dimensions ($p < .001$). For the 2nd task Ss were presented the original stimuli and others which differed systematically from them. Ss stated whether or not they had seen each stimulus during Task 1. Ss who responded to certain dimensions during Task 1 identified stimuli with new values of those dimensions better than their controls ($p < .001$). It was concluded that cues may acquire distinctiveness if the cue is considered to be a value of a stimulus dimension and not a stimulus configuration per se.

In this study an attempt was made to demonstrate acquired distinctiveness of cues (ADC) using human *S*s and a paired-associate learning task. Robinson (1955), and Hake and Eriksen (1955, 1956) attempted to demonstrate ADC using this method; but their studies did not successfully do so. There is good reason to believe that ADC is a demonstrable phenomenon since Lawrence (1949, 1950) demonstrated it using discrimination tasks and rats as *S*s. This study is similar to those of Lawrence in that the cues in ADC refer to stimulus dimensions. Cues were considered to be stimulus configurations in the Robinson, and Hake and Eriksen studies.

Two tasks were utilized in the present study. First *S*s learned different verbal responses to certain designs, and second they were asked to identify, from a larger group of designs, those of the first task. It is hypothesized that if the cues acquired distinctiveness during the first task certain of the designs would be easier to identify for some *S*s than for others. It is also hypothesized that if *S*s were responding to the stimulus aspects rather than the designs qua configurations there would be differences in the learning rates using the same stimuli and the same responses.

METHOD

Materials.—The materials which were used were designs drawn on 3 × 5 index cards. Each design was a triangle with a little cross drawn somewhere on the card. By using different shape triangles and making them different sizes and putting the cross in different positions in relation to the triangle the design could be varied in three formally independent dimensions. By using 3 values of each of these dimensions, that is, 3 shapes, 3 sizes, and 3 positions, there are 27 possible designs which can be drawn. The first task used 18 of these designs.

For the second task all of the possible combinations (27 designs) were used. Twenty other designs were also used here. These included 6 which consisted of one of the aforementioned shapes and sizes but with the cross

---

[1] This paper is based on a PhD dissertation submitted to the Graduate School of the University of Minnesota. The author is indebted to James J. Jenkins for his advice and encouragement during the course of this study. A portion of this study was read at the Midwestern Psychological Association Meeting in Chicago, May 1963.

[2] Now at Arlington State College.

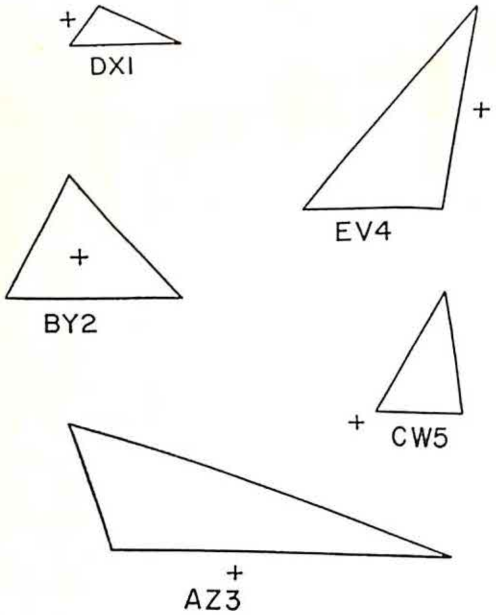


FIG. 1. Representative designs shown to *S*s during the two tasks of the experiment. (The first letter of the name is associated with the shape; the second with the size; and the number is associated with the position of the cross. Forty-seven combinations of these dimensions are used.)

in one of two new positions (position stimuli), and 6 with one of the aforementioned sizes and positions with one of two new shapes (shape stimuli), and 6 with one of the aforementioned shapes and positions and one of two new sizes (size stimuli). The other 2 designs had a new size, shape, and position (all aspects new stimuli).

All shapes, sizes, and positions are represented in Fig. 1. The designs can be identified if the shapes are labeled A, B, C, D, and E; the sizes labeled V, W, X, Y, and Z; and the positions labeled 1, 2, 3, 4, and 5. Thus any design can be identified by the label for its shape, size, and position, e.g., A-Y-1.

Procedure.—Each *S* was assigned to one of seven experimental groups and given two tasks to perform. The first task was a paired-associate learning task with 18 designs used as stimuli. The verbal responses varied from group to group.

Groups I, II, and III had 3 different responses; Groups IV and V had 9 different responses, and Groups VI and VII had 18 different responses to learn to the 18 stimuli. The responses for Groups I, II, and III were the verbal "D-1," "F-6," and "G-3." These were associated with shape values in Group I, size values in Group II, and position value in Group III. The dimensions not associated were irrelevant.

The responses for Groups IV and V had two independent elements: the first was one of the three letters A, B, and C; and the second was one of the three numbers 1, 2, and 3. Independent combinations of these give nine responses, e.g., "B-2." The letters related to the shapes for Groups IV and V, and the numbers related to the size in Group IV, and position in Group V.

For Groups VI and VII a new element was added to the responses of Groups IV and V. One of the letters X, Y, or Z was combined with A, B, or C and 1, 2, or 3 for one of 27 possible responses, e.g., "A-Y-3." However, if A, B, and C corresponded to the shapes; X, Y, and Z to sizes; and 1, 2, and 3 to positions, each stimulus was uniquely identified with one of 18 different responses. These pairings were made for Group VI.

The stimuli presented to Group VII were given the same responses as those of Group VI; however, the responses were randomly related to the stimulus configurations.

Each *S* was given 90 presentations of the stimuli, or five randomizations of the 18 stimuli. The *S* had the possible responses on a card before him. After each response by *S* *E* gave the correct response for that stimulus, or if the response were acceptable, *E* said "correct." Comparisons were made between the groups on the number of correct responses to the stimuli to test for differences in the rate of learning.

Approximately 1 min. after Task 1, Task 2 was run. The second set of cards was presented one at a time to *S*. For these he responded "yes" if he thought it was different from all of the cards in Task 1, and "no" if he thought it was the same as at least one of the cards in Task 1. Forty-seven cards were randomly presented three times each.

ADC was tested by comparing the different groups' responses to the different stimulus classes by analyses of variance. The 18 learning stimuli, the 9 filler stimuli, the 6 position stimuli, the 6 size stimuli, the 6 shape stimuli, and the 2 all-aspects-new stimuli composed the stimulus classes compared.

Subjects.—The *S*s were 152 volunteers from the elementary psychology courses at the University of Minnesota. Four of these were discarded due to errors of procedure. Eight *S*s (7 from Group II, 1 from Group I) failed to reach a learning criterion (three *SD*s above chance) which was established for Groups I, II, and III; and were not used in the analyses of the data of Task 2.

TABLE 1

COMPARISONS OF NUMBER OF CORRECT RESPONSES IN TASK 1 USING t TESTS BETWEEN ADJACENT GROUPS

	III		I		II		V		IV		VI		VII
	(pos)		(sh)		(si)		(sh-pos)		(sh-si)		(sh-si-pos)		(random)
$\bar{X}$	87.45		72.85		62.74		55.90		38.65		20.55		3.65
df		19		46		45		33		32		19	
t		4.35***		1.77		1.12		2.65**		2.77***		4.76***	

** $p < .01$.
*** $p < .001$.

RESULTS

An overall analysis of variance was not performed on the number of correct responses during the learning task due to great heterogeneity of variance. The largest variance was 359 times as large as the smallest. However, a median test showed significant differences in central tendency, χ^2 (6) = 73.3, $p < .001$.

The seven groups were arranged in order by mean number of correct responses. Adjacent groups were compared by t tests. The results are given in Table 1. It can be seen that there are significant differences between groups with the same number of responses, e.g., Groups I, II, and III, as well as between groups with different numbers of responses. Generally, the specific dimensions which were used for discrimination determined the ease of learning the discrimination, as did the number of cues which had to be discriminated and the number of different responses.

The recognition task (Task 2) showed that certain of the cues were more discriminable than others, irrespective of the specific cues to which the responses were learned during the first task. Table 2 illustrates this result. The all-aspects-new stimuli were most easily identified. Second most easily identified were the position stimuli. Next in ease of identification were the shape stimuli. There were, however, no differences in the number of "no" responses to the size stimuli, the filler stimuli, and the learning stimuli. The number of "no" responses which are recorded in Table 2 are "correct" for all stimuli except the learning stimuli.

There were between-group differences as well as between-stimuli

TABLE 2

MEAN NUMBER OF TIMES PER 18 RESPONSES *S*s IDENTIFIED DIFFERENT CLASSES OF STIMULI OF THE SECOND TASK AS DIFFERENT FROM ANY STIMULUS OF THE FIRST TASK

Stimulus Class	I	II	III	IV	V	VI	VII
	(sh)	(si)	(pos)	(sh-si)	(sh-pos)	(sh-si-pos)	(random)
All-aspects-new	14.1	12.9	17.40	16.20	17.25	16.95	15.20
Position	11.05	10.30	17.40	13.05	16.25	16.30	13.60
Shape	7.00	4.05	2.30	9.75	9.55	9.10	4.80
Size	1.15	2.50	1.30	2.75	.75	2.55	1.45
Filler	1.87	2.37	1.23	2.60	.77	3.23	1.40
Learning	1.93	1.88	1.52	2.27	.63	2.72	1.68

differences. The *S*s in the groups who learned to respond to the position of the cross identified significantly more stimuli with new values of this dimension than did the other groups. A contrast of these two sets of groups was highly significant, F (1, 133) = 43.18, $p < .001$. The same kind of an analysis gave virtually identical results for the number of identifications of the new values of the shape dimension. This time *S*s who learned to respond to the shape identified a significantly greater number of stimuli, F (1, 133) = 41.71, $p < .001$. The same results also held for the size stimuli, F (1, 133) = 16.50, $p < .001$; however, in this case the results were confounded by these *S*s also identifying a greater number of the stimuli to which the responses were learned in Task 1 as being "new." When this tendency was statistically controlled these differences disappeared, F (6,132) = 1.619, $p > .05$.

DISCUSSION

The *S*s who learned to give responses based upon values of certain dimensions of stimulus configurations identified new values of those dimensions better than their controls. This remained true whether *S* responded to values of one, two, or three dimensions of the stimulus. Furthermore, the accuracy of this identification was positively related to the original saliency of the dimension discriminated. The accuracy also was relatively independent of simple verbal labels of the dimension values, since the values of the most salient dimension were not easily labeled. These results are interpreted as evidence for ADC. The *S*s who learned to give specific responses to certain "cues" later discriminated those cues better than *S*s who did not respond specifically to those cues. This interpretation is valid only if one considers the cues to be values of stimulus dimensions and not stimulus configurations qua configurations.

REFERENCES

HAKE, H. W., & ERIKSEN, C. W. Effect of number of permissible response categories on learning of a constant number of visual stimuli. *J. exp. Psychol.*, 1955, **50**, 161–167.

HAKE, H. W., & ERIKSEN, C. W. Role of response variables in recognition and identification of complex visual forms. *J. exp. Psychol.*, 1956, **52**, 235–243.

LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimuli. *J. exp. Psychol.*, 1949, **39**, 770–784.

LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.

ROBINSON, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, 1955, **49**, 112–115.

(Received July 27, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 591–592

SUPPLEMENTARY REPORTS

NONREINFORCEMENTS OF PERCEPTUAL AND MEDIATING RESPONSES IN CONCEPT LEARNING[1]

HOWARD H. KENDLER
University of California, Santa Barbara

AND

MARGARET WOERNER
New York University

After college students learned to classify geometrical designs into 1 of 2 categories with a chain of behavior involving an initial perceptual-orienting response, a mediating response, and a final-choice response, they received a series of 10 nonreinforcements followed by either a reversal or nonreversal shift. The perceptual-orienting response of half of the Ss during the extinction series was not followed by the appearance of the designs. For the remaining Ss it was. Ss who had not been shown the designs exhibited the customary superiority of a reversal over a nonreversal shift. The other Ss did not, presumably because their mediational response had been extinguished.

The purpose of this experiment was to determine the effects a series of nonreinforcements had upon a concept-learning chain that had already been acquired. The chain contained an initial perceptual-orienting response (looking to the left or right), followed in order by a mediational reaction and a terminal choice response (pressing one of two keys). A light signal, following the correct choice, had served as the reinforcement for the entire chain.

The results of a previous study (Kendler, Glucksberg, & Keston, 1961) suggested that the perceptual and mediational segments were relatively independent, that is, the mediational response was not strongly associated with the cues resulting from the orienting response but instead was connected primarily to the geometrical designs that had to be classified. If this analysis is correct, a series of nonreinforcements should have different effects depending on whether or not the perceptual-orienting response led to the appearance of the geometrical designs. If it did, the superiority of a reversal shift over a nonreversal shift should be reduced or eliminated because the mediational response, presumably responsible for this superiority (Kendler & Kendler, 1962), would be weakened by the nonreinforcements. On the other hand if the perceptual-orienting response did not lead to the appearance of the geometrical design, no opportunity for the extinction of the mediational response would occur. Under such conditions the superiority of a reversal over a nonreversal shift should be maintained.

Method.—The experimental procedure was essentially similar to that used by Kendler et al. (1961). The *S* sat close to a translucent screen through which were projected for 5 msec. two geometrical figures, one at each side. The *S* had to learn to pay attention to the set of figures on one side only and sort them into one of two categories represented by two response keys. The major procedural modification instituted in the present study was a series of interpolated nonreinforced trials (a light indicating a correct choice never flashed on) between the learning of the two successive concepts. After learning the first concept all *S*s (a total of 48 students from an introductory psychology course) received 10 trials in which no reinforcement was forthcoming regardless of their response. For half of the *S*s (PE) a blank slide was projected on the screen. When *S* looked to one side he would see no geometrical figure and therefore it was assumed that the perceptual-orienting segment of the previously successful chain was weakened through nonreinforcement. The remaining half of the *S*s (ME) were exposed to the same set of geometrical designs which they categorized during the learning of the initial concept. Since no reinforcement was obtained for their choice behavior, presumably their mediational response, as well as their orienting response, was being extinguished. Half of the PE and ME *S*s received a reversal shift (R) following the interpolated trials while the other half were given a nonreversal shift (NR).

The relevant stimulus figure remained on the same side of the screen before and after the shift. Any *S* who failed to respond during one of the 10 interpolated trials was reminded to

---

[1] This study was supported by the Office of Naval Research under Contract Nonr-285(30) with New York University.

TABLE 1

Mean and Median Number of Trials to Learn Second Concept

Group	*M*	*SD*	Mdn.	Range
PE:R	8.8	11.3	2.5	1–43
PE:NR	31.2	37.3	19.5	0–128
ME:R	19.2	16.5	18.0	1–43
ME:NR	23.1	34.4	11.0	2–128
R[a]	4.7	4.9	2.5	1–18
NR[a]	41.4	46.0	17.5	3–128

[a] Data from Kendler et al. (1961).

press one of the two response keys during every trial.

Results and discussion.—The four groups did not differ significantly in the speed in which they learned the first concept. Table 1 reports the learning scores for the second (postshift) concept. Results of equivalent reversal and nonreversal groups *without* interpolated trials (Kendler et al., 1961) are included in the lower two lines.

The data are consistent with the expectation that failing to reinforce the perceptual-orienting response in the absence of the stimuli that evoke mediational responses would not disturb the commonly observed superiority of a reversal over a nonreversal shift. The difference between the performance of Groups PE:R and PE:NR was significant in the direction predicted at the .025 level (Mann & Whitney, 1947). In contrast there was no significant difference between the reversal and nonreversal shifts when the mediational extinction procedure was used, that is, between Groups ME:R and ME:NR.

A comparison of the present results with those of *S*s who performed without interpolated nonreinforcement trials *suggests* that the mediational extinction procedure retards a reversal shift while facilitating a nonreversal shift, although the differences between Groups R and ME:R, and NR and ME:NR failed to achieve statistical significance.

These results once again demonstrate that for human adults the dominant mode of response to the present kind of problem is to execute a reversal shift. When the implicit response that mediates the reversal shift is weakened by nonreinforcements the reversal tendency loses its dominant position in the response hierarchy. The results also show that an essential segment of a problem-solving chain, in this case the perceptual-orienting response, can be weakly integrated with the succeeding links of the chain as evidenced by the fact that failing to reinforce it does not disturb the dominance of a reversal over a nonreversal shift.

REFERENCES

Kendler, H. H., Glucksberg, S., & Keston, R. Perception and mediation in concept learning. *J. exp. Psychol.*, 1961, **61**, 186–191.

Kendler, H. H., & Kendler, T. S. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, **69**, 1–16.

Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **17**, 50–60.

(Received June 24, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 592–594

A REPLICATION OF PAIRED-ASSOCIATE LEARNING AS A FUNCTION OF S-R SIMILARITY[1]

SLATER E. NEWMAN
North Carolina State College

This experiment replicated the finding that paired-associate learning is interfered with when the same items occur as S terms and R terms (Umemoto & Hilgard, 1961; Young, 1961). It demonstrated also that learning is interfered with when S terms share with R terms all the same letters but no items.

Method.—Seventy-two female undergraduates at Meredith College served as *S*s. They were assigned to one of six groups in the order in which they appeared at the laboratory. None had served before in a verbal-learning experiment.

Three types of six-pair lists were used: one in which the S terms (CVC trigrams) and R terms were identical, but with identical terms never paired (the Identical or I list); one in which the same nine consonants used in con-

---

[1] This research was supported by Contract Nonr 486(08) between the Office of Naval Research and North Carolina State College. This is Technical Report No. 8 prepared under this contract. The contribution of Linda Marsh Green in gathering the data and performing some of the analyses is gratefully acknowledged. Reproduction of this article in whole or in part is permitted for any purpose of the United States government.

TABLE 1

MEANS AND SIGNIFICANCE TESTS DURING TRAINING, R-S RECALL, AND S-TERM RECALL

Group	S-R Training				R-S Errors	R-S/Sub. (%)	R-S Recall				S-Term Recall
	Correct	Inven.	Omis.	Substi.[c]			Correct	Inven.	Omis.	Substi.	Correct
Identical (I)	15.63	6.17	86.42	17.79	6.38	.358	.67	.29	4.13	.92	—
Similar (S)	27.00	11.67	80.92	6.42	1.38	.214	1.00	.71	4.13	.17	1.79
Dissimilar (D)	56.04	5.42	59.96	4.58	1.17	.255	2.08	.54	3.29	.08	2.92
I vs. S	5.06[a]**	1.90[b]	1.23[a]	3.46[b]**	4.38[b]**		1.18[a]	1.70[b]	0[a]	2.88[a]**	
D vs. S	4.59[b]**	2.10[b]*	4.01[a]**	1.17[a]	.50[a]		2.34[b]*	.96[b]	.91[a]	.27[b]	2.81[a]**

[a] *t* test.
[b] *z* based on Mann-Whitney test.
[c] R-S plus non-R-S substitutions.
* $p < .05$.
** $p < .01$.

structing the R terms were used for the S terms (the Similar or S list); and one in which the nine consonants used with the S terms differed from those for the R terms (the dissimilar or D list). No two S terms of a list began with the same letter or ended with the same letter. The same was true for R terms. For S terms and R terms, each of the five vowels and Y were used as middle letters.

Two sets of I, S, and D lists were constructed. The three lists of a set used the same R terms. The association values (Archer, 1960) of the six lists were approximately equivalent.

S-term similarity was the same for all six lists; so was R-term similarity. Only S-R similarity varied, the I and S lists differing in the number of terms (6 vs. 0) which were both S and R terms and thus in the number of different terms used in the list (6 vs. 12); the D and S groups differing in the number of letters (6 vs. 18) shared by S and R terms and thus in the number of different letters used in the list (24 vs. 15).

Each *S* was exposed to one of the lists for 21 trials. Each of the 21 pairing trials was followed by a test trial. On pairing trials each of the six pairs was presented; on test trials each of the S terms was presented alone, *S*'s task being to spell out the R term. A 2-sec. rate was used on pairing and test trials. The intertrial interval was 4 sec. Three different pairing orders and three different test orders were used to reduce the likelihood that serial position would serve as a cue. For each of the three lists with the same set of R terms, the sequence of R terms for the three pairing-trial orders, and of their correct emission on the three test-trial orders was the same.

After the twenty-first test, each *S* was given an R-S test. Each R term was presented for 2 sec., and *S* was instructed to give its S term. S-term recall followed for the S and D groups. The *S* was given 45 sec. to call out all the S terms he could remember.

Results.—Since the results for each dependent variable when Set A served as the R terms were substantially the same as those for the lists with Set B as R terms, the data for the two replications of each treatment have been combined. The means for each dependent variable are given in Table 1.

t tests, or where variance was heterogeneous, Mann-Whitney tests were done. Separate tests were used to compare the S group with the I and the D groups on each variable. The results of these tests are presented in Table 1.

The S group made more correct responses during paired-associate training than the I group ($p < .01$), and fewer substitution errors ($p < .01$). There were no differences ($p > .05$) in omissions or inventions.

The number of substitutions which were R-S errors was determined for the I group by counting the number of times each term was given as a response to the term for which it was the stimulus (e.g., ZAF-HUX was one pair, GIM-ZAF another. An R-S error was counted if *S* gave GIM instead of HUX when ZAF was presented). The extent to which the same substitutions (e.g., GIM instead of HUX) occurred in the S group was also determined. The *S*s in the I group made more of these errors ($p < .01$), and these formed a higher percentage of their substitutions than in the S group. The I group also made more non-R-S substitutions ($p < .05$) than the S group.

Eight of the S group and five of the I group gave each of the R terms at least once during training, χ^2 (1) = .94, $p > .05$. Associative learning was studied by determining for each *S* the number of pairs for which the first time he gave an R term he gave it to its S term. The number of *S*s in each group who gave the R term correctly the first time for more than half the R terms given was counted. There were 6 in the I group and 15 in the S group, χ^2 (1) = 6.86, $p < .01$. On the R-S test, the only difference between the two groups was in substitutions; the I group made more ($p < .01$).

The D group made more correct responses during paired-associate training than the S group ($p < .01$) but fewer inventions ($p < .05$) and omissions ($p < .01$). No differences were observed in substitutions ($p > .05$). Fifteen of the D group gave each of the R terms at least once during training, and only 8 of the S group χ^2 (1) = 4.08, $p < .05$. Eighteen of the D group and 15 of the S group gave more than half the R terms correctly the first time they gave them, χ^2 (1) = .88, $p > .05$.

On the R-S test, the D group made more correct responses ($p < .05$) than the S group. There were no differences in omissions, inventions, or substitutions ($p > .05$). The D group made more correct responses in S-term recall ($p < .01$).

Discussion.—When the same item served as both an S and R term, learning was slower than when it served as one or the other. This replicates previous findings (Umemoto & Hilgard, 1961; Young, 1961). Substitutions, both R-S and others increased suggesting that the increased interference was not just a function of R-S associations. The increase in substitutions may have been partly due to *S*'s failure to discriminate between S and R terms early during training (Umemoto & Hilgard, 1961).

The comparison of S and D groups indicated that paired-associate learning and R-S recall are inversely related to number of letters shared between S and R terms. This accords with findings from experiments in which formal similarity has been varied among S terms or R terms (Feldman & Underwood, 1957; Horowitz, 1962; Newman & Buckhout, 1962) or between them (Young, 1961). Paired-associate learning appears to be inversely related to the number of elements (items or letters) shared between S and R terms.

REFERENCES

ARCHER, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

FELDMAN, S., & UNDERWOOD, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53**, 11–15.

HOROWITZ, L. M. Associative matching and intralist similarity. *Psychol. Rep.*, 1962, **10**, 751–757.

NEWMAN, S. E., & BUCKHOUT, R. S-R and R-S learning as functions of intralist similarity. *Amer. J. Psychol.*, 1962, **75**, 429–436.

UMEMOTO, T., & HILGARD, E. R. Paired-associate learning as a function of similarity: Common stimulus and response items within the list. *J. exp. Psychol.*, 1961, **62**, 97–104.

YOUNG, R. K. Paired-associate learning when the same items occur as stimuli and responses. *J. exp. Psychol.*, 1961, **61**, 315–318.

(Received July 1, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 594–595

TRANSFER FROM SERIAL TO PAIRED-ASSOCIATE LEARNING[1]

ROBERT K. YOUNG AND MICHAEL CASEY
University of Texas

Transfer from a serial to a paired-associate (PA) list was investigated. In the experimental condition both serial and PA lists contained the same items, arranged in the same order. In the control condition the same PA list was preceded by a serial list composed of different items. If associations are formed between the items of a serial list, the experimental PA list should be learned faster than its control. No differences between the learning of the 2 PA lists were obtained. The results of the present experiment were interpreted as being unfavorable to an inter-item association theory of serial learning.

Horowitz and Izawa (1963) have recently investigated transfer from serial to paired-associate (PA) learning as a function of associative relationships among the items employed in the lists. In their experiment, the arrangement of items in the serial list was identical to the arrangement of the same items in the subsequent PA list. Varying amounts of positive transfer were obtained in the various conditions of their experiment. The transfer measure used was based on a comparison of the learning of the PA list preceded

---

[1] The present study was supported in part by a grant from the Graduate School of the University of Texas.

by the serial list with the learning of the same PA list *not* preceded by the serial list.

Young (1961) has assumed that if associations are formed between the adjacent items of a serial list, positive transfer should be obtained in the learning of a subsequent PA list, the pairs of which are formed of adjacent items from the serial list. The finding of positive transfer by Horowitz and Izawa would appear to be strong evidence for a theory which holds that associations are formed between items in a serial list.

The results of the Horowitz and Izawa experiment do not, however, permit an unambiguous interpretation. The transfer measure used was based on a comparison of a group which had previously learned a serial list with one which had not. The obtained transfer might be a function of interitem associations formed during serial learning, or it might be the result of differential practice. Since either of these would be expected to produce positive transfer, the obtained transfer cannot be conclusively attributed to the formation of interitem associations.

The present experiment is designed to control for differential practice effects in the investigation of transfer from serial to PA learning. To control for the possibility that Young's (1961) failure to obtain positive transfer in a similar experiment was due to the use of a select group of adjectives, the items used in the present study were taken from the condition of the Horowitz and Izawa experiment which yielded the greatest positive transfer.

Method.—Two 10-item serial lists were constructed, with the first item of each serving as the anticipation symbol. Two 9-pair double-function PA lists were also constructed, one from each of the serial lists. The items used in the serial and PA lists as well as their orders and pairings were taken from the first 20 items of Cond. I of the Horowitz and Izawa experiment. As in that experiment, each item in the PA lists (except the first and last) served as the stimulus in one pair and as the response in another.

A total of 48 *S*s, randomly selected from introductory psychology classes, was divided into two groups. The first group learned one serial list and the second learned the other serial list. Each group was then subdivided, each half learning one of the two PA lists. Three *E*s were used yielding a $2 \times 2 \times 3$ factorial design ($N = 4$ per cell).

All lists were learned by the anticipation method to a criterion of one perfect recitation. Approximately 5 min. separated the learning of the serial and PA lists to permit changing lists and giving PA instructions. The serial list was presented at a 2-sec. rate with a 6-sec. intertrial interval and the PA list was presented at a 2:2-sec. rate with a 4-sec. intertrial interval. Both lists were presented on a Lafayette memory drum. Three orders of presentation of the PA lists were used to avoid serial learning. Two *S*s were dropped because of failure to follow instructions and each was replaced by the next *S* to appear.

Results and discussion.—The mean numbers of trials to learn the two serial lists were found to be 11.63 and 11.71, $F\ (1, 36) < 1.00$, for List I and List II, respectively. The lists may be considered to be of equivalent difficulty.

Trials to learn the PA list were also analyzed. PA learning following relevant serial learning (i.e., serial lists formed of the same items as the PA lists) did not differ from PA learning following irrelevant serial learning, $F\ (1, 36) < 1.00$. The difference was in the direction predicted by the associative hypothesis, the mean numbers of trials to learn the PA list following relevant and irrelevant serial learning being 23.71 and 25.21, respectively.

Learning of individual items in the PA list as a function of position held in the previous serial list was also analyzed. In contrast to previous experiments (Young, 1961, 1962), no differences were observed for individual items in the PA list as a function of prior relevant or irrelevant serial learning.

The results of the present study do not support a theory of interitem associations formed during serial learning. They do replicate the results of previous investigations of transfer from serial to PA learning (Young, 1961, 1962). The present results also suggest that the positive transfer observed by Horowitz and Izawa may have resulted from practice effects. Their results, therefore, cannot be taken as unambiguously supporting an interitem association theory of serial learning.

REFERENCES

Horowitz, L. W., & Izawa, C. Comparison of serial and paired-associate learning. *J. exp. Psychol.*, 1963, **65**, 352–361.

Young, R. K. The stimulus in serial verbal learning. *Amer. J. Psychol.*, 1961, **74**, 517–528.

Young, R. K. Tests of three hypotheses about the effective stimulus in serial learning. *J. exp. Psychol.*, 1962, **63**, 307–313.

(Received July 15, 1963)

Journal of Experimental Psychology
1964, Vol. 67, No. 6, 596

SPEED VERSUS MINIMUM-CHOICE INSTRUCTIONS IN CONCEPT ATTAINMENT

PATRICK R. LAUGHLIN

Northwestern University

To remove ambiguity in the instructions in previous concept-attainment research, 2 groups each containing 16 college students solved 5 concept-attainment problems. A speed group was instructed to attain the concepts as quickly as possible, regardless of number of card choices; a minimum-choice group was instructed to attain the concepts in as few choices as possible, regardless of time. The 2 groups did not differ in number of card choices, while the speed group attained the concepts significantly faster.

In a series of experiments Bruner, Goodnow, and Austin (1956) studied selection strategies in concept attainment. These experiments involved "efficiency instructions," in which S was instructed "to arrive at the concept as efficiently as possible [p. 83]." It would seem that these instructions could be interpreted by S as to attain the concept (*a*) in as few choices as possible, regardless of time, (*b*) as quickly as possible, regardless of number of choices, (*c*) some undetermined weighting of both. The present study explicitly separated these possible interpretations through the instructions given to S.

Method.—The procedure of Bruner, Goodnow, and Austin (1956) was followed with geometric form problems of six attributes and two values of each. One group of male college students ($N = 16$) was instructed to attain the concept in as few choices as possible, regardless of time; a second group ($N = 16$) was instructed to attain the concept as quickly as possible, regardless of the number of card choices. The *S*s were randomly assigned to the two groups. Half of the *S*s in each group had an ordered problem board, in which the position of each card varied systematically in relation to the other cards (for example, all cards with one border were in the top four rows, all cards with two borders in the bottom four rows), half a randomly arranged board, in which the card position did not vary systematically. Five problems, having one, two, three, two, and one relevant attributes, respectively, were used for all *S*s.

Results and discussion.—The mean number of card choices to solution for the five problems combined was 41.12 for the minimum-choice group and 44.75 for the speed group, with *SD*s of 14.12 and 16.23, respectively. An overall analysis of variance showed no significant difference in number of card choices ($F < 1$). The mean time to solution for the five problems combined was 20.18 min. for the minimum-choice group and 11.01 min. for the speed group, with *SD*s of 10.60 and 4.72, respectively. Overall analysis of variance showed a significant difference in favor of the speed group in time to solution, $F(1, 30) = 9.38$, $p < .005$. There were no differences for the ordered vs. random problem boards.

These results indicate that minimum-choice and speed instructions should be explicitly differentiated in future concept-attainment research. They would further seem to indicate a motivational effect of speed instructions on performance, since the two groups differed in speed of solution but not in accuracy or efficiency as measured by the number of card choices.

REFERENCE

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

(Received July 11, 1963)

NEW FROM SCIENTIFIC PROTOTYPE

THE NEW LDR SWITCH

that provides transient free audio switching with the lowest distortion ever offered.

THE NEW MODEL GB TACHISTOSCOPE

All new with many improvements over the world famous model GA. 3-channel. Color corrected optical system. Direct drive lamp circuits. Many other unique features.

THE NEW REED-RELAY SHOCK SCRAMBLER

Absolutely truly random shock scrambler with low power consumption and high reliability.

NEW CATALOGS . . . GET YOUR COPIES

Psychobiological Equipment.
Digital Test Equipment.

SCIENTIFIC PROTOTYPE MFG. CORP.
623 W. 129th St. **New York 27, N.Y.**

Behavior Analysis Aparatus . . . operant conditioning kits, rodent test cages, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, timers, switches, digital test equipment and special systems.

ZEITSCHRIFT FÜR EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Herausgeber für Deutschland
PROF. DR. PH. LERSCH, München PROF. DR. F. SANDER, Bonn
PROF. DR. K. GOTTSCHALDT, Göttingen

Herausgeber für Österreich
PROF. DR. H. ROHRACHER, Wien

Herausgeber für die Schweiz
PROF. DR. R. MEILI, Bern

Redaktion
DR. C. J. HOGREFE, Göttingen

Band X/4

THEO HERRMANN, **Untersuchungen zum Erlernen von Wahlreaktionen**

WALTER KRISTOF, **Eine Faktorenanalyse geometrisch-optischer Täuschungen**

DIETER V. KLEBELSBERG, **Ergebnis einer Verhaltensanalyse des Kraftfahrers**

KLAUS FOPPA, RAINER FUCHS UND HEINZ HECKHAUSEN, **Tübinger Symposion über Motivation**

KRITISCHE BUCHBESPRECHUNG
HANS THOMAE, **Günter Blau und Elisabeth Müller-Luckmann: Gerichtliche Psychologie**

All articles are provided with summaries in English and French.
The journal appears quarterly and contains 160 pages per number.

Subscription is DM 64,—per Volume

VERLAG FÜR PSYCHOLOGIE—DR. C. J. HOGREFE—GÖTTINGEN
(Brentanoweg 10, Germany)